UNIVERSIDAD DEL CEMA
Buenos Aires
Argentina

Serie
DOCUMENTOS DE TRABAJO

Área: Matemática, Economía, Negocios, Ingeniería

Serie de Machine Learning
Revisión de Algebra Lineal 1

Sergio A. Pernice

Julio 2020
Nro. 736

www.cema.edu.ar/publicaciones/doc_trabajo.html
UCEMA: Av. Córdoba 374, C1054AAP Buenos Aires, Argentina
ISSN 1668-4575 (impreso), ISSN 1668-4583 (en línea)
Editor: Jorge M. Streb; asistente editorial: Valeria Dowding jae@cema.edu.ar

Serie de Machine Learning
Revisión de Algebra Lineal 1

SERGIO A. PERNICE[1]

Universidad del CEMA
Av. Córdoba 374, Buenos Aires, 1054, Argentina

16 de Julio de 2020

Abstract

En este documento presentamos una primera revision de álgebra lineal de una forma especialmente adaptada para sus eventuales aplicaciones en aprendizaje automático (machine learning). Es el primero de una serie de documentos sobre machine learning en español. Es parte del contenido del curso "Métodos de Machine Learning para Economistas" de la Maestría en Economía de la UCEMA.

Keywords: álgebra lineal, regresiones, machine learning, aprendizaje automático.

Contents

[1] sp@ucema.edu.ar
Los puntos de vista del autor no representan necesariamente la posición de la Universidad del CEMA.

1 Introducción

En esta introducción se asume conocido el material presentado en la sección 1.1.

1.1 Notación y conceptos básicos

Usamos letras minúsculas en negrita "$\mathbf{v}$" para vectores, letras minúsculas "a" para escalares, y mayúsculas "A" para matrices. A menos que se especifique lo contrario los vectores son "columna", y si queremos referirnos a vectores "fila" lo hacemos explícitamente con el símbolo de traspuesto $\mathbf{v}^\top$.

Vector columna $\mathbf{x} \in \mathbb{R}^n$:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \tag{1.1}$$

Una matriz $A \in \mathbb{R}^{m\times n}$ se puede pensar simplemente como una tabla de $m \times n$ entradas, como n vectores columna, cada uno $\in \mathbb{R}^{m\times 1}$, o m vectores fila, cada uno $\in \mathbb{R}^{1\times n}$:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} | & | & & | \\ a^1 & a^2 & \cdots & a^n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} -a_1^\top- \\ \vdots \\ -a_m^\top- \end{bmatrix} \tag{1.2}$$

La matriz identidad es I, y sus elementos son δ_{ij}:

$$I_{ij} = \delta_{ij} = \begin{cases} 1, & i = j; \\ 0, & i \neq j. \end{cases} \tag{1.3}$$

$IA = AI = A$, $I\mathbf{v} = \mathbf{v}$, $\mathbf{v}^\top I = \mathbf{v}^\top$.

Una matriz diagonal tiene todos los elementos no diagonales iguales a cero: $D_{ij} = 0$ si $i \neq j$. $D = \mathrm{diag}(d_1, d_2, \cdots, d_n)$:

$$D_{ij} = \begin{cases} d_i, & i = j; \\ 0, & i \neq j. \end{cases} \tag{1.4}$$

Obviamente I es una matriz diagonal donde $d_i = 1, \forall i$.

Dependiencia/independencia lineal: Se dice que un conjunto de vectores $\{\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_n\} \subset \mathbb{R}^m$ es linealmente dependiente si un vector perteneciente al conjunto se puede representar como una combinación lineal de los vectores restantes; es decir,

$$\mathbf{x}_n = \sum_{i=1}^{n-1} \alpha_i \mathbf{x}_i \tag{1.5}$$

para algún conjunto de escalares $\alpha_1, \ldots, \alpha_{n-1} \in \mathbb{R}$. Si no existe un vector para el cual exista tal conjunto de escalares, entonces el conjunto se dice linealmente independiente.

Si el conjunto es linealmente independiente, cualquier vector en el "span" (conjunto de todas las posibles combinaciones lineales de ellos) de $\{\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_n\} \subset \mathbb{R}^m$ se puede expresar de manera única como combinación lineal de ellos, mientras que si es linealmente dependiente hay infinitas maneras de hacerlo (dada una combinación lineal cuyo resultado sea el vector deseado, siempre le podemos sumar el vector $\mathbf{0}$, como por ejemplo pasando en (1.5) el lado derecho al izquierdo cambiando el signo y multiplicando luego por un escalar arbitrario α).

Para el **producto escalar (o "producto interno")** entre vectores vamos a usar la forma matricial:

$$\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^\top \mathbf{y} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} x_i y_i \in \mathbb{R} \tag{1.6}$$

Con el producto escalar se puede definir la **"norma", o magnitud, o longitud de vectores** como

$$||\mathbf{x}||_2 = \sqrt{\mathbf{x}^\top \mathbf{x}} \tag{1.7}$$

y la **distancia entre dos vectores x e y** como $||\mathbf{x} - \mathbf{y}||$.

Ángulo entre vectores: Con el producto escalar (1.6), el ángulo θ entre dos vectores $\mathbf{x}$ e $\mathbf{y}$ viene dado por

$$\cos\theta = \frac{\mathbf{x}^\top \mathbf{y}}{||\mathbf{x}||_2 ||\mathbf{y}||_2} \tag{1.8}$$

los vectores $x/||\mathbf{x}||_2$ tienen norma 1, y si $\cos\theta = 0$ se dice que los vectores son ortogonales (o perpendiculares, el ángulo entre ellos es de $\pi/2$ o $3\pi/2$).

Si $\mathbf{x}$ tiene norma 1, $\mathbf{x}^\top \mathbf{y}$ es la **proyección** de $\mathbf{y}$ en la dirección de $\mathbf{x}$, y $(\mathbf{x}^\top \mathbf{y})\,\mathbf{x}$ es un vector que apunta en la dirección de $\mathbf{x}$ y su norma es la proyección de $\mathbf{y}$ en la dirección de $\mathbf{x}$.

1.2 Producto "outer" (o outer product)

El **producto "outer"** entre dos vectores $\mathbf{x} \in \mathbb{R}^m$ e $\mathbf{y} \in \mathbb{R}^n$ es:

$$\mathbf{x} \otimes \mathbf{y} \equiv \mathbf{x}\, \mathbf{y}^\top = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix} = \begin{bmatrix} x_1 y_1 & x_1 y_2 & \cdots & x_1 y_n \\ x_2 y_1 & x_2 y_2 & \cdots & x_2 y_n \\ \vdots & \vdots & \ddots & \vdots \\ x_m y_1 & x_m y_2 & \cdots & x_m y_n \end{bmatrix} \in \mathbb{R}^{m \times n} \tag{1.9}$$

Es decir, $\mathbf{x} \otimes \mathbf{y}$ es el producto matricial entre $\mathbf{x} \in \mathbb{R}^{m\times 1}$ e $\mathbf{y}^\top \in \mathbb{R}^{1\times n}$. Las dimensiones de este producto son: $(m \times 1)(1 \times n) \to (m \times n)$. Notar que la componente ij de $\mathbf{x} \otimes \mathbf{y}$ es $x_i y_j$.

Veamos el efecto de multiplicar por derecha un vector columna por el producto outer entre dos vectores (asumimos que las dimensiones son correctas):

$$(\mathbf{x} \otimes \mathbf{y})\, \mathbf{z} = (\mathbf{x}\, \mathbf{y}^\top)\, \mathbf{z} = \mathbf{x}\, (\mathbf{y}^\top \mathbf{z}) \tag{1.10}$$

Pensado como producto matricial, el producto outer satisface la propiedad asociativa. La última igualdad, sin embargo, tiene gran utilidad, porque el producto escalar entre $\mathbf{y}$ y $\mathbf{z}$ va a ser cero si esos vectores son ortogonales. Es decir que el efecto de la matriz $\mathbf{x}\ \mathbf{y}^\top$ sobre el vector $\mathbf{z}$ es multiplicar *escalarmente* a $\mathbf{z}$ por $\mathbf{y}^\top$ y al escalar resultante multiplicarlo por el vector $\mathbf{x}$. El resultado tiene la dirección de $\mathbf{x}$. Por lo tanto si $\mathbf{y} = \mathbf{x}$ y $|\mathbf{x}| = 1$, $\mathbf{x}\ \mathbf{x}^\top$ es un proyector en el subespacio spaneado por $\mathbf{x}$.

De la misma manera, veamos ahora el efecto de multiplicar por izquierda un vector fila por el producto outer entre dos vectores (nuevamente asumimos que las dimensiones son correctas):

$$\mathbf{z}^\top\, (\mathbf{x} \otimes \mathbf{y}) = \mathbf{z}^\top\, (\mathbf{x}\, \mathbf{y}^\top) = (\mathbf{z}^\top \mathbf{x})\, \mathbf{y}^\top \tag{1.11}$$

Además de la propiedad asociativa por derecha y por izquierda, el producto outer satisface las siguientes propiedades:

$$\begin{aligned} (\mathbf{x} \otimes \mathbf{y})^\top &= (\mathbf{y} \otimes \mathbf{x}) \\ (\mathbf{y} + \mathbf{z}) \otimes \mathbf{x} &= \mathbf{y} \otimes \mathbf{x} + \mathbf{z} \otimes \mathbf{x} \\ \mathbf{x} \otimes (\mathbf{y} + \mathbf{z}) &= \mathbf{x} \otimes \mathbf{y} + \mathbf{x} \otimes \mathbf{z} \\ c(\mathbf{x} \otimes \mathbf{y}) &= (c\mathbf{x}) \otimes \mathbf{y} = \mathbf{x} \otimes (c\mathbf{y}) \end{aligned} \tag{1.12}$$

1.3 Diferentes maneras de ver los productos matriciales

Diferentes maneras de ver el producto "matrix-vector":

$$\mathbf{y} = A\mathbf{x} = \begin{bmatrix} -a_1^\top- \\ \vdots \\ -a_m^\top- \end{bmatrix} \mathbf{x} = \begin{bmatrix} a_1^\top \mathbf{x} \\ \vdots \\ a_m^\top \mathbf{x} \end{bmatrix} \tag{1.13}$$

cada componente es el producto escalar del correspondiente vector fila de A con $\mathbf{x}$. Enfatiza la relación de ángulo entre los vectores fila de A y el vector $\mathbf{x}$. Otra manera es:

$$\mathbf{y} = A\mathbf{x} = \begin{bmatrix} | & & | \\ a^1 & \cdots & a^n \\ | & & | \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} | \\ a^1 \\ | \end{bmatrix} x_1 + \cdots + \begin{bmatrix} | \\ a^n \\ | \end{bmatrix} x_n \tag{1.14}$$

el resultado es una combinación lineal de los vectores columna de A, enfatiza el span de estos vectores.

(1.14) es especialmente útil, entre otras cosas, para entender cuándo una ecuación lineal $A\mathbf{x} = \mathbf{b}$ va a tener solución, con la matriz A y el vector $\mathbf{b}$ conocidos, siendo el vector $\mathbf{x}$ la incógnita. De acuerdo a (1.14), $A\mathbf{x}$, para todos los posibles vectores $\mathbf{x}$, es el span de los vectores columna de A, por lo que la ecuación va a tener solución sólo si $\mathbf{b}$ está en dicho span. Por otro lado si las columnas de A son linealmente independientes y $\mathbf{b}$ está en su span, la solución es única, mientras que si las columnas son linealmente dependientes y $\mathbf{b}$ está en su span, van a haber infinitas soluciones.

Ejercicio: verificar (1.13) y (1.14).

En componentes,

$$y_i = \sum_{j=1}^{n} a_{ij} x_j \tag{1.15}$$

si pensamos en a_{ij} como la componente j de un vector i tenemos (1.13). Si la miramos como la componente i de un vector j tenemos (1.14).

Diferentes maneras de ver el producto "vector-matrix":

Pensando en la matriz como conjunto de vectores columna:

$$\mathbf{y}^\top = \mathbf{x}^\top A = \mathbf{x}^\top \begin{bmatrix} | & | & & | \\ a^1 & a^2 & \cdots & a^n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} \mathbf{x}^\top a^1 & \mathbf{x}^\top a^2 & \cdots & \mathbf{x}^\top a^n \end{bmatrix} \tag{1.16}$$

enfatiza la relación de ángulo entre el vector $\mathbf{x}^\top$ y las columnas de A.

Pensando en la matriz como conjunto de vectores fila:

$$\mathbf{y}^\top = \mathbf{x}^\top A = \begin{bmatrix} x_1 & \cdots & x_m \end{bmatrix} \begin{bmatrix} -a_1^\top- \\ \vdots \\ -a_m^\top- \end{bmatrix} = x_1 \left[-a_1^\top-\right] + \ldots + x_m \left[-a_m^\top-\right] \tag{1.17}$$

el producto es una combinación lineal de estos vectores. Enfatiza el span de los vectores fila de A.

Ejercicio: verificar (1.16) y (1.17).

En componentes,

$$y_j = \sum_{i=1}^{n} x_i a_{ij} \tag{1.18}$$

si pensamos en a_{ij} como la componente i de un vector j tenemos (1.16). Si la miramos como la componente j de un vector transpuesto i tenemos (1.17).

Diferentes manera de ver el producto "matriz-matriz":

Como un conjunto de productos escalares:

$$C = AB = \begin{bmatrix} -a_1^\top- \\ \vdots \\ -a_m^\top- \end{bmatrix} \begin{bmatrix} | & & | \\ b^1 & \cdots & b^p \\ | & & | \end{bmatrix} = \begin{bmatrix} a_1^\top b^1 & a_1^\top b^2 & \cdots & a_1^\top b^p \\ a_2^\top b^1 & a_2^\top b^2 & \cdots & a_2^\top b^p \\ \vdots & \vdots & \ddots & \vdots \\ a_m^\top b^1 & a_m^\top b^2 & \cdots & a_m^\top b^p \end{bmatrix} \tag{1.19}$$

Como conjunto de productos “matriz-vector”:

$$C = AB = A \begin{bmatrix} | & & | \\ b^1 & \cdots & b^p \\ | & & | \end{bmatrix} = \begin{bmatrix} | & & | \\ Ab^1 & \cdots & Ab^p \\ | & & | \end{bmatrix} \tag{1.20}$$

Como conjunto de productos “vector-matriz”:

$$C = AB = \begin{bmatrix} -a_1^\top- \\ \vdots \\ -a_m^\top- \end{bmatrix} B = \begin{bmatrix} -a_1^\top B- \\ \vdots \\ -a_m^\top B- \end{bmatrix} \tag{1.21}$$

Como suma de productos outer:

$$C = AB = \begin{bmatrix} | & & | \\ a^1 & \cdots & a^n \\ | & & | \end{bmatrix} \begin{bmatrix} -b_1^\top- \\ \vdots \\ -b_n^\top- \end{bmatrix} = \sum_{i=1}^{n} \mathbf{a}^i \, \mathbf{b}_i^\top \tag{1.22}$$

Ejercicio: verificar (1.19-1.22): A: $m \times n$, B: $n \times p$, C: $m \times p$.

Las compenentes de $C = AB$ son:

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} \tag{1.23}$$

Si pensamos a a_{ik} como la componente k del vector fila i, y a b_{kj} como la componente k del vector columna j tenemos (1.19).

Si pensamos a b_{kj} como la componente k del vector columna j y a a_{ik} como el elemento ik de la matriz A, tenemos (1.20).

Si pensamos a a_{ik} como la componente k del vector fila i, y a b_{kj} como el elemento kj de la matriz B, tenemos (1.21).

Finalmente, si pensamos a a_{ik} como la componente i del vector columna k y a b_{kj} como la componente j del vector fila k, entonces tenemos (1.22).

Propiedades de la multiplicación de matrices:

1. Asociativa: $(AB)C = A(BC)$
2. Distributiva: $A(B + C) = AB + AC$
3. En general no es conmutativa: $AB \neq BA$ en general (puede ocurrir que un producto exista y el otro no, pero aun si ambos existen puede que no sean iguales).

1.4 La transpuesta de una matriz

Ya la estuvimos utilizando, pero resumamos aquí sus propiedades. Si el elemento ij de A son a_{ij}, el elemento ij de $A^\top$ son a_{ji}.

1. $(A^\top)^\top = A$
2. $(AB)^\top = B^\top A^\top$
3. $(A+B)^\top = A^\top + B^\top$

1.5 La traza de una matriz

$A \in \mathbb{R}^{n\times n}$ (cuadrada) con elementos a_{ij}:

$$\mathrm{tr} A = \sum_{i=1}^{n} a_{ii} \tag{1.24}$$

A menos que se explicite de otra manera, las siguientes propiedades asumen todas las matrices involucradas $\in \mathbb{R}^{n\times n}$:

1. $\mathrm{tr} A = \mathrm{tr} A^\top$
2. $\mathrm{tr}(A+B) = \mathrm{tr} A + \mathrm{tr} B$
3. $\mathrm{tr}(\alpha A) = \alpha \mathrm{tr}(A)$
4. A: $m\times n$, B: $n \times m$, $\mathrm{tr}(AB) = \mathrm{tr}(BA)$:

$$\mathrm{tr}(AB) = \sum_{i=1}^{m}\left(\sum_{j=1}^{n} a_{ij}b_{ji}\right) = \sum_{j=1}^{n}\left(\sum_{i=1}^{m} b_{ji}a_{ij}\right) = \mathrm{tr}(BA)$$

5. Para A, B, C tal que ABC es cuadrada, $\mathrm{tr}\, ABC = \mathrm{tr}\, BCA = \mathrm{tr}\, CAB$, y lo mismo vale para productos de mas matrices (sale de 4 y propiedad asociativa).

1.6 La noción de norma

La norma de un vector $\mathbf{x}$, informalmente, es la "longitud" de un vector. Ya conocemos la norma Euclidiana (1.7) (o norma ℓ_2, como vamos a definir ahora): $||\mathbf{x}||_2^2 = \mathbf{x}^\top\mathbf{x}$, pero la idea se puede generalizar.

Una "norma" es cualquier función: $\mathbb{R}^n \to \mathbb{R}$ tal que satisface las siguientes 4 propiedades:

1. $\forall \mathbf{x} \in \mathbb{R}^n, f(\mathbf{x}) \geq 0$ (no-negatividad)

2. $f(\mathbf{x}) = 0 \Leftrightarrow \mathbf{x} = \mathbf{0}$

3. $\forall \mathbf{x} \in \mathbb{R}^n, t \in \mathbb{R}, f(t\mathbf{x}) = |t| f(\mathbf{x})$ (homogeneidad).

4. $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n, f(\mathbf{x} + \mathbf{y}) \leq f(\mathbf{x}) + f(\mathbf{y})$ (desigualdad triangular)

Ejemplos de normas:

$\ell_2 : \ ||\mathbf{x}||_2 = \mathbf{x}^\top \mathbf{x} = \sqrt{\sum_{i=1}^n x_i^2}$

$\ell_1 : \ ||\mathbf{x}||_1 = \sum_{i=1}^n |x_i|$

$\ell_\infty : \ ||\mathbf{x}||_\infty = \max_i |x_i|$

$\ell_p : \ ||\mathbf{x}||_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$

Ejercicio: probar que $\ell_\infty = \lim_{p \to \infty} \ell_p$.

Todas estas normas cumplen con las 4 propiedades y pueden ser útiles en algunos problemas.

Normas de una matriz: hay muchas, la más simple es considerarla como un vector, se conoce como la norma de Frobenius:

$$||A||_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2} = \sqrt{\operatorname{tr}(A^\top A)} \tag{1.25}$$

2 Funciones lineales y rango de una matriz

2.1 Matrices como funciones

El **rango columna** de una matriz es el máximo número de columnas linealmente independientes. El **rango fila** es el máximo número de filas linealmente independientes. Resulta ser que para toda matriz, el rango fila es igual al rango columna.

Este resultado se puede probar en dos líneas, pero en general, esas dos líneas no proveen la intuición necesaria para que el resultado sea completamente transparente, ya que es una propiedad que no es intuitivamente obvia al principio, no al menos de la manera en que sí lo son otros resultados en álgebra lineal. Esto es desafortunado, ya que es un resultado importante. Y es importante que sea completamente transparente, porque varios otros resultados dependen de este, y la mente humana no es buena para digerir resultados que dependen de otros resultados si esos otros no resultan obvios. Por eso tomamos una ruta más larga para explicar este resultado. El objetivo, precisamente, es que al final de la ruta el resultado finalmente resulte obvio.

Matrices como vectores:

Antes que nada enfaticemos, como ya lo hemos hecho, que las matrices son un conjunto de números que se pueden pensar de muchas maneras diferentes. Por ejemplo, una matriz $A \in \mathbb{R}^{m\times n}$ puede pensarse como un vector $\in \mathbb{R}^{mn\times 1}$:

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \longrightarrow \begin{pmatrix} a_{11} \\ a_{21} \\ a_{12} \\ a_{22} \end{pmatrix} \tag{2.1}$$

la transformación (2.1) es lineal. La norma de Frobenius (1.25) de la matriz no es otra cosa que la norma ℓ_2 si la miramos como vector.

Matrices como funciones lineales

Para el propósito en consideración es especialmente útil mirar a las matrices como representantes de funciones lineales que a un vector le asignan otro vector. Es decir, $A \in \mathbb{R}^{m\times n}$ como representante a una función lineal $f_A \colon \mathbb{R}^{n\times 1} \to \mathbb{R}^{m\times 1}$, como en $A\mathbf{x} = \mathbf{y}$, y/o como representante de otra función lineal $g_A \colon \mathbb{R}^{1\times m} \to \mathbb{R}^{1\times n}$, como en $\mathbf{y}^\top A = \mathbf{x}^\top$.

Para ello conviene primero ver propiedades de las funciones lineales en abstracto. La Fig. 1 nos recuerda las nociones de **dominio**, **codominio** e **imagen** o **rango** de una función.

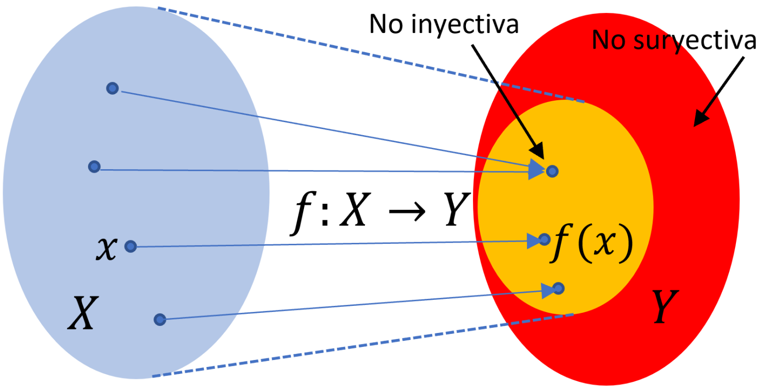


Figure 1: X, en celeste, es el **dominio** de la función f. f aplica a todos los elementos $x \in X$. Y, en rojo, es el **codominio** de f. No todo elemento de Y es "receptor" de la función f. El conjunto anaranjado, de elementos $y \in Y$ para los que existe un $x \in X$ tal que $f(x) = y$, se conoce como la **imagen** o **rango** de f.

Funciones inyectivas, suryectiva y biyectiva

Una función es **inyectiva** si asigna elementos distintos de su dominio a elementos distintos de su codominio. En otras palabras, cada elemento del codominio de la función es la imagen de, como máximo, un elemento de su dominio. La función de la Fig 1 claramente no es inyectiva ya que hay un elemento de la imagen al que le llegan 2 flechas.

Una función es **suryectiva** si para cada elemento y en el codominio Y hay al menos un elemento x en el dominio X tal que $f(x) = y$. No se requiere que x sea único; la función puede asignar uno o más elementos de X al mismo elemento de Y. La Fig 1 sugiere que la función representada no es inyectiva ya que a la parte roja de Y (que presumiblemente es $\neq \emptyset$) no le llega ninguna flecha.

Una función es **biyectiva** (o "uno a uno") si a cada elemento del codominio le corresponde exactamente por un elemento del dominio y vice versa. Es decir, la función es tanto inyectiva como suryectiva.

2.2 Funciones lineales, formulación abstracta

Si V y W son espacios vectoriales reales de dimensiones n y m respectivamente, una función $f : V \to W$ es **lineal** si

$$\mathbf{f}(a\mathbf{x} + b\mathbf{y}) = a\mathbf{f}(\mathbf{x}) + b\mathbf{f}(\mathbf{y}) \tag{2.2}$$

donde $\mathbf{x}, \mathbf{y} \in V$, $\mathbf{f}(\mathbf{x}), \mathbf{f}(\mathbf{y}) \in W$, y $a, b \in \mathbb{R}$.

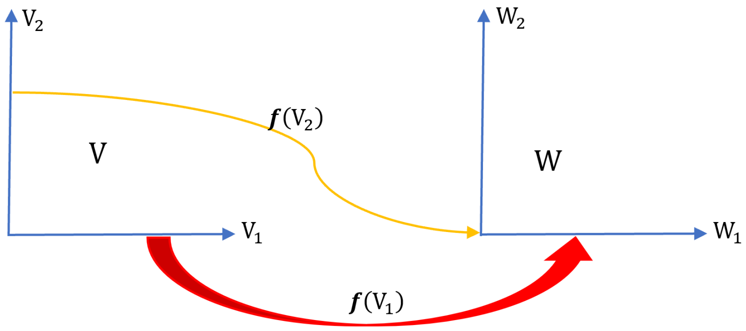


Figure 2: $\dim V = \dim V_1 + \dim V_2$, and $\dim V_1 = \dim W_1$.

Consideremos el conjunto $V_2 \subset V$ tal que si $\mathbf{v}_2 \subset V_2$, $\mathbf{f}(\mathbf{v}_2) = \mathbf{0}$, ver figura 2. V_2 es un subespacio vectorial de V que se conoce como espacio nulo de $\mathbf{f}$, o $\mathcal{N}(\mathbf{f})$.

Para ver esto tomemos dos vectores $\mathbf{v}_{2,a}$ y $\mathbf{v}_{2,b} \in V_2$ ($\mathbf{f}(\mathbf{v}_{2,a}) = \mathbf{f}(\mathbf{v}_{2,b}) = \mathbf{0}$) no nulos, linealmente independientes, y dos escalares α_a y α_b, al menos uno de ellos no nulo. Entonces, por linealidad, $\mathbf{f}(\alpha_a \mathbf{v}_{2,a} + \alpha_b \mathbf{v}_{2,b}) = \mathbf{0}$, por lo que $\alpha_a \mathbf{v}_{2,a} + \alpha_b \mathbf{v}_{2,b} \in V_2$. ✓

Ahora consideremos el *complemento ortogonal* de V_2 en V, al que llamamos V_1 ($V_1 = V_2^{\perp}$). V_1 es un subespacio de V, es decir, es cerrado frente a operaciones lineales: si $\mathbf{v}_{1,a}$ y $\mathbf{v}_{1,b} \in V_1$, para todo elemento de $\mathbf{v}_2 \in V_2$, $\mathbf{v}_2^{\top}\mathbf{v}_{1,a} = \mathbf{v}_2^{\top}\mathbf{v}_{1,b} = 0$, entonces, por linealidad, $\mathbf{v}_2^{\top}(\alpha_a \mathbf{v}_{1,a} + \alpha_b \mathbf{v}_{1,b}) = 0$. Además, para todo elemento no nulo $\mathbf{v}_1 \in V_1$, $\mathbf{f}(\mathbf{v}_1) \neq \mathbf{0}$, ya que si fuera $\mathbf{0}$ por definición estaría en V_2.

Sin embargo notar que la inversa no vale, que $\mathbf{v}$ sea tal que $\mathbf{f}(\mathbf{v}) \neq \mathbf{0}$ no implica que $\mathbf{v} \in V_1$, ya que, por ejemplo, si $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$, donde $\mathbf{v}_1 \in V_1$ y $\mathbf{v}_2 \in V_2$, $\mathbf{f}(\mathbf{v}) = \mathbf{f}(\mathbf{v}_1) \neq \mathbf{0}$ y sin embargo, por construcción, $\mathbf{v} \notin V_1$.

Todo elemento $\mathbf{v} \in V$ se puede expresar de manera única de la forma $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$, donde $\mathbf{v}_1 \in V_1$ y $\mathbf{v}_2 \in V_2$ (se dice que V es la "suma directa" de V_1 y V_2, $V = V_1 \oplus V_2$, $V_1 \cap V_2 = \{\mathbf{0}\}$), y $\mathbf{v}_1^\top \mathbf{v}_2 = 0$.

Ya sabemos que la imagen de V_2 es $\{\mathbf{0}\}$, consideremos ahora la imagen de V_1, a la que llamamos $W_1 \subset W$, ver figura 2, es decir,

$$W_1 = \{\mathbf{w} \in W \ / \ \exists \ \mathbf{v} \in V_1, \mathbf{f}(\mathbf{v}) = \mathbf{w}\} \tag{2.3}$$

Resulta que $\dim W_1 = \dim V_1$.

Para ver esto simplemente consideremos una base de vectores linealmente independientes de V_1, $\{\mathbf{x}_1, \ldots, \mathbf{x}_d\}$, $d \leq n$, y elijamos los dos primeros, $\mathbf{x}_1$ y $\mathbf{x}_1$. $\mathbf{f}(\mathbf{x}_1)$ y $\mathbf{f}(\mathbf{x}_2)$ son linealmente independientes. Si no lo fueran, existirían dos escalares α_1 y α_2, con al menos uno de ellos no nulo, tales que $\alpha_1 \mathbf{f}(\mathbf{x}_1) + \alpha_2 \mathbf{f}(\mathbf{x}_2) = \mathbf{0}$. Pero $\alpha_1 \mathbf{f}(\mathbf{x}_1) + \alpha_2 \mathbf{f}(\mathbf{x}_2) = \mathbf{f}(\alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2)$, por lo que, o bien $\alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2 = \mathbf{0}$, o bien $\in V_2$. Pero ninguna de esas alternativas es posible, ya que por hipótesis $\mathbf{x}_1$ y $\mathbf{x}_2$ son linealmente independientes y V_1 es cerrado ante operaciones lineales.

Consideremos ahora cualquier vector del span de $\{\mathbf{x}_1, \mathbf{x}_2\}$, al que llamamos $\mathbf{x}_{12}$ y $\mathbf{x}_3$. Por el mismo razonamiento que antes, $\mathbf{f}(\mathbf{x}_{12})$ y $\mathbf{f}(\mathbf{x}_3)$ son linealmente independientes. Iterando hasta completar la base $\{\mathbf{x}_1, \ldots, \mathbf{x}_d\}$ demostramos que el conjunto de vectores $\{\mathbf{f}(\mathbf{x}_{12}), \ldots, \mathbf{f}(\mathbf{x}_3)\}$ es linealmente independiente. Hay j elementos en este conjunto, y para cualquier $\mathbf{x}$ que no pertenezca al span de la base $\{\mathbf{x}_1, \ldots, \mathbf{x}_d\}$, como $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$, con $\mathbf{x}_1 \in V_1$ y $\mathbf{x}_2 \in V_2$, $\mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{x}_1)$, por lo tanto esta en el span de $\{\mathbf{f}(\mathbf{x}_{12}), \ldots, \mathbf{f}(\mathbf{x}_3)\}$. Por lo tanto $\dim W_1 = \dim V_1$. ✓

Acabamos de ver, de manera abstracta, que $n = \dim V = \dim V_1 + \dim V_2 = j + (n - j)$, y que $j = \dim V_1 = \dim W_1$. W_2 no tiene ninguna restricción inducida por $\mathbf{f}$, aunque, obviamente, si $\dim W = m$, $\dim W_2 = m - j$. W es el codominio de $\mathbf{f}$ y W_1 su imagen o rango.

Otra manera de decir que lo probamos más arriba es que f es una biyección entre V_1 y W_1 que respeta las operaciones lineales.

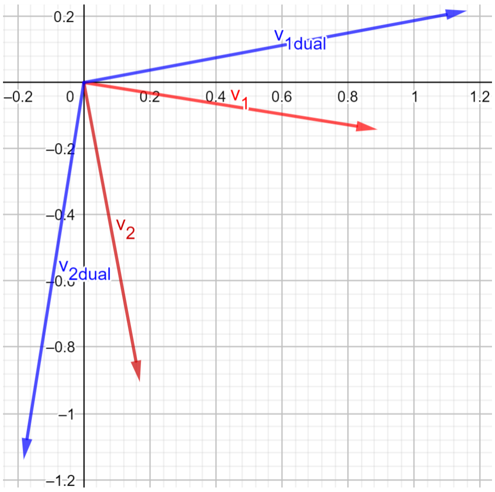
Las funciones lineales quedan determinadas por su efecto en una base del dominio, como lo muestra el siguiente teorema.

Teorema: Supongamos que $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ es una base de V no necesariamente linealmente independiente, y $\mathbf{f}\colon\ V \to W$ es lineal. $\mathbf{f}$ queda definida por su imagen en estos vectores $\{\mathbf{f}(\mathbf{x}_1), \mathbf{f}(\mathbf{x}_2), \ldots, \mathbf{f}(\mathbf{x}_n)\}$.

Demostración: ya que $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ es una base de V, todo vector $\mathbf{z}$ de V se puede escribir como combinacion lineal de ellos $\mathbf{z} = \sum_{i=1}^n a_i \mathbf{x}_i$, y (2.2) indica que $\mathbf{f}(\mathbf{z}) = \sum_{i=1}^n a_i \mathbf{f}(\mathbf{x}_i)$. ✓

2.3 Base dual

Para conectar la idea abstracta de una función lineal entre espacios vectoriales reales con su representación como matrices conviene antes definir la **base dual** de una base dada, ver que tal base siempre existe, y entender cómo construirla.



(a) Base $\{\mathbf{v}_1, \mathbf{v}_2\}$ (roja) y su base dual $\{\mathbf{v}^1, \mathbf{v}^2\}$ (azul).

```
np.random.seed(111321)
# c1 y c2 son dos vectores 2x1 con elementos aleatorios
v1 = 2*(np.random.rand(2,1)-0.5)
v2 = 2*(np.random.rand(2,1)-0.5)

# vector en la direccion dual de v1 (ortogonal a v2)
v1d = scipy.linalg.null_space(np.transpose(v2))
# vector en la direccion dual de v2 (ortogonal a v1)
v2d = scipy.linalg.null_space(np.transpose(v1))

v1d /= np.dot(np.transpose(v1),v1d) # vector dual de v1
v2d /= np.dot(np.transpose(v2),v2d)  # vector dual de v2

print("v1 =",v1)
print("v2 =",v2)
print("v1d =",v1d)
print("v2d =",v2d)

print("v1.v1d =", np.dot(np.transpose(v1),v1d))
print("v1.v2d =", np.dot(np.transpose(v1),v2d))
print("v2.v2d =", np.dot(np.transpose(v2),v2d))
print("v2.v1d =", np.dot(np.transpose(v2),v1d))
```

```
v1 = [[ 0.88887704]
 [-0.14203138]]
v2 = [[ 0.16938612]
 [-0.90469276]]
v1d = [[1.15971014]
 [0.21713316]]
v2d = [[-0.1820676 ]
 [-1.13943625]]
v1.v1d = [[1.]]
v1.v2d = [[-2.77555756e-17]]
v2.v2d = [[1.]]
v2.v1d = [[2.49800181e-16]]
```

(b) Python function “scipy.linalg.null space()”.

Figure 3: Dada cualquier base de vectores linealmente independientes $\{\mathbf{v}_1, \mathbf{v}_2\}$, siempre podemos encontrar la base dual $\{\mathbf{v}^1, \mathbf{v}^2\}$ tal que $\left(\mathbf{v}^i\right)^\top \mathbf{v}_k = \delta_{i,k}$.

Dada la base $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ de $\mathbb{R}^{n\times 1}$, siempre podemos encontrar otra base de $\mathbb{R}^{n\times 1}$, que llamamos “dual” de $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$, a la denotamos con super-índices en vez de sub-índices, $\{\mathbf{v}^1, \ldots, \mathbf{v}^n\}$, tal que para todo i

$$\left(\mathbf{v}^i\right)^\top \mathbf{v}_k = \delta_{i,k} \tag{2.4}$$

ver figura 3.

En tratamientos clásicos de álgebra lineal, las bases duales se plantean generalmente como bases en el espacio de funcionales lineales $f\colon V \to \mathbb{R}$ (o el field sobre el que este definido el espacio V). En este documento preferimos enfatizar el sentido geométrico de las bases duales, por lo que las definimos como bases del mismo V que satisfacen los productos escalares (2.4). En dimensión finita, que es el caso en este documento, ambos tratamientos son equivalentes (equivalencia que se conoce como “Riesz representation theorem”).

Veamos intuitivamente que esa base existe y cómo se puede construir: para cada i, los vectores $\{\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}, \mathbf{v}_{i+1} \ldots \mathbf{v}_n\}$ (es decir, la base completa *con exclusión de* $\mathbf{v}_i$) spanean un subespacio de $n-1$ dimensiones. Llamamos $\tilde{\mathbf{v}}^i$ a un vector genérico no-nulo en el complemento ortogonal de dicho subespacio, es decir $\tilde{\mathbf{v}}^i$ es tal que

$$\left(\tilde{\mathbf{v}}^i\right)^\top \mathbf{v}_k = 0, \quad \mathbf{v}_k \in \{\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}, \mathbf{v}_{i+1} \ldots \mathbf{v}_n\} \tag{2.5}$$

Obviamente tiene que ocurrir que

$$\left(\tilde{\mathbf{v}}^i\right)^\top \mathbf{v}_i = c \neq 0 \tag{2.6}$$

ya que $\mathbf{v}_i$ tiene que tener una componente no nula en la dirección de $\tilde{\mathbf{v}}^i$, de lo contrario $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ (con *inclusión* de $\mathbf{v}_i$), no sería una base de $\mathbb{R}^{n\times 1}$ como habíamos supuesto. El vector

$$\mathbf{v}^i = \frac{\tilde{\mathbf{v}}^i}{c} \tag{2.7}$$

satisface (2.4). Ver la figura 3b y el uso de la función **scipy.linalag.null space** para ver como encontrar $\tilde{\mathbf{v}}^i$, a partir de una matriz cuyas columnas son los vectores $\{\mathbf{v}_1, \dots, \mathbf{v}_{i-1}, \mathbf{v}_{i+1} \dots \mathbf{v}_n\}$ (notar que hay que transponer dicha matriz).

Es importante notar que la base dual de $\{\mathbf{v}^1, \dots, \mathbf{v}^n\}$ es $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$, es decir, la base original. Y que si dicha base es ortonormal, la base "dual" coincide con la original.

Por último, supongamos que el espacio vectorial V es la suma directa de dos subespacios ortogonales entre sí, V_1 y V_2, como en la figura 2. Consideremos la base

$$\left\{\mathbf{v}_{1,1}, \dots, \mathbf{v}_{1,j}, \mathbf{v}_{2,j+1} \dots \mathbf{v}_{2,n}\right\} \tag{2.8}$$

donde $\left\{\mathbf{v}_{1,1}, \dots, \mathbf{v}_{1,j}\right\}$ es base de V_1 y $\left\{\mathbf{v}_{2,j+1} \dots \mathbf{v}_{2,n}\right\}$ es base de V_2. Como V_1 y V_2 son ortogonales entre sí,

$$\mathbf{v}_{1,i}^{\top}\mathbf{v}_{2,k} = 0, \forall i = 1, \dots j, k = j+1, \dots n \tag{2.9}$$

en la base dual de (2.8), $\left\{\mathbf{v}^{1,1}, \dots, \mathbf{v}^{1,j}, \mathbf{v}^{2,j+1} \dots \mathbf{v}^{2,n}\right\}$,

$$\left\{\mathbf{v}^{1,1}, \dots, \mathbf{v}^{1,j}\right\} \text{ es base de } V_1 \tag{2.10}$$

y

$$\left\{\mathbf{v}^{2,j+1} \dots \mathbf{v}^{2,n}\right\} \text{ es base de } V_2 \tag{2.11}$$

Es decir si $V = V_1 \oplus V_2$, los subespacios V_1 y V_2 son cerrados respecto de la operación de dualidad de una base dada.

Para ver esto basta considerar un elemento de (2.10): $\mathbf{v}^{1,i}$, $i = 1, \dots j$. (2.4) indica que $\left(\mathbf{v}^{1,i}\right)^{\top} \mathbf{v}_{2,k} = 0, \forall k = j+1, \dots n$, por lo que $\mathbf{v}^{1,i} \notin V_2$, y como $V_1 \cap V_2 = \{\mathbf{0}\}$ y $\mathbf{v}^{1,i} \neq \mathbf{0}$, $\mathbf{v}^{1,i} \in V_1$.

De la misma manera, consideremos un elemento de (2.11): $\mathbf{v}^{2,k}$, $k = j+1, \dots n$. (2.4) indica que $\left(\mathbf{v}^{2,k}\right)^{\top} \mathbf{v}_{1,i} = 0, \forall i = 1, \dots j$, por lo que $\mathbf{v}^{2,k} \notin V_1$, y como $V_1 \cap V_2 = \{\mathbf{0}\}$ y $\mathbf{v}^{2,k} \neq \mathbf{0}$, $\mathbf{v}^{2,k} \in V_2$. ✓

2.4 Conexión entre funciones lineales y matrices

Teorema: hay una relación 1 a 1 entre funciones lineales $\mathbf{f}_A\colon V \to W$ y matrices $A\colon \in \mathbb{R}^{\dim W \times \dim V}$.

2.4.1 Dada una función lineal $\mathbf{f}_A$, encontrar la matriz A correspondiente

Supongamos que $\dim V = n$ y $\dim W = m$. Primero mapeamos los espacios vectoriales abstractos V y W en $V \leftrightarrow \mathbb{R}^{n\times 1}$ y $W \leftrightarrow \mathbb{R}^{m\times 1}$ (es decir, columnas de n y m números reales respectivamente), que son "realizaciones" de dichos espacios.

Elijamos n vectores columna de $\mathbb{R}^{n\times 1}$ linealmente independientes, y los ordenamos así:

$$\{\mathbf{v}_{1,1},\ldots,\mathbf{v}_{1,j},\mathbf{v}_{2,j+1},\ldots,\mathbf{v}_{2,n}\}\subset\mathbb{R}^{n\times 1},\quad j\leq n \tag{2.12}$$

De la misma manera, elijamos m vectores columna de $\mathbb{R}^{m\times 1}$ linealmente independientes, y los ordenamos así.

$$\{\mathbf{w}_{1,1},\ldots,\mathbf{w}_{1,j},\mathbf{w}_{2,j+1}\ldots,\mathbf{w}_{2,m}\}\subset\mathbb{R}^{m\times 1},\quad j\leq m \tag{2.13}$$

Consideremos la función lineal $f\colon\mathbb{R}^{n\times 1}\to\mathbb{R}^{m\times 1}$ definida por:

$$\mathbf{f}(\mathbf{v}_{1,i}) = \mathbf{w}_{1,i},\quad i=1,\ldots,j \tag{2.14}$$

$$\mathbf{f}(\mathbf{v}_{2,k}) = \mathbf{0},\quad k=j+1,\ldots,n \tag{2.15}$$

como vimos antes, (2.14-2.15) definen completamente a la función.

Ahora consideremos la matriz

$$A_i=\mathbf{w}_{1,i}\otimes\mathbf{v}^{1,i}=\mathbf{w}_{1,i}\left(\mathbf{v}^{1,i}\right)^{\top} \tag{2.16}$$

donde estamos utilizando el producto outer definido en la sección 1.2 y los vectores $\mathbf{v}^{1,i}$ pertenecen a la base dual de (2.12). (2.4) indica que

$$A_i\mathbf{v}_{1,i} = \mathbf{w}_{1,i} \tag{2.17}$$

$$A_i\mathbf{v}_{a,k} = \mathbf{0}\quad \text{si } a=2 \text{ o si } k\neq i \tag{2.18}$$

La matriz A correspondiente a la función lineal (2.14-2.15) es:

$$A=\sum_{i=1}^{j}A_i=\sum_{i=1}^{j}\mathbf{w}_{1,i}\left(\mathbf{v}^{1,i}\right)^{\top} \tag{2.19}$$

Notar que si bien en la fórmula (2.19) no aparecen explícitamente los vectores $\mathbf{v}^{2,i}$, ya que $0^{m\times 1}\otimes\mathbf{v}_{2,i}=0^{m\times n}$, su presencia influye en la construcción de la base "dual" de (2.12)

$$\{\mathbf{v}^{1,1},\ldots,\mathbf{v}^{1,j},\mathbf{v}^{2,j+1},\ldots,\mathbf{v}^{2,n}\} \tag{2.20}$$

ya que todos los $\{\mathbf{v}^{1,1},\ldots,\mathbf{v}^{1,j}\}$ tienen que ser ortogonales a todos los vectores $\{\mathbf{v}^{2,j+1},\ldots,\mathbf{v}^{2,n}\}$, como indican (2.10) y (2.11). ✓

2.4.2 Dada una matriz A, encontrar la función lineal $\mathbf{f}_A$ correspondiente: el problema

Dada una matriz $A\in\mathbb{R}^{m\times n}$, queremos determinar dos bases

$$\{\mathbf{v}_{1,1},\ldots,\mathbf{v}_{1,j},\mathbf{v}_{2,j+1},\ldots,\mathbf{v}_{2,n}\}\subset\mathbb{R}^{n\times 1},\quad j\leq n \tag{2.21}$$

$$\{\mathbf{w}_{1,1},\ldots,\mathbf{w}_{1,j},\mathbf{w}_{2,j+1},\ldots,\mathbf{w}_{2,m}\}\subset\mathbb{R}^{m\times 1},\quad j\leq m \tag{2.22}$$

tal que:

$$A\mathbf{v}_{1,i} = \mathbf{w}_{1,i} \tag{2.23}$$
$$A\mathbf{v}_{1,i} = \mathbf{0} \tag{2.24}$$

de modo que la función $\mathbf{f}_A\colon \mathbb{R}^{n\times 1} \to \mathbb{R}^{m\times 1}$ quedará definida por:

$$\mathbf{f}\left(\mathbf{v}_{1,i}\right) = \mathbf{w}_{1,i}, \quad i = 1, \ldots, j \tag{2.25}$$
$$\mathbf{f}\left(\mathbf{v}_{2,k}\right) = \mathbf{0}, \quad k = j+1, \ldots, n \tag{2.26}$$

Antes de ver la solución general del problema planteado en (2.21-2.26), veamos un ejemplo.

2.4.3 Ejemplo

Consideremos la matriz

$$A = \begin{pmatrix} 1 & 0 & a \\ 0 & 1 & b \end{pmatrix} \in \mathbb{R}^{2\times 3} \tag{2.27}$$

Notemos que

$$\begin{pmatrix} x & 0 & 0 \\ y & 0 & 0 \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \end{pmatrix} \tag{2.28}$$
$$\begin{pmatrix} 0 & x & 0 \\ 0 & y & 0 \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}\begin{pmatrix} 0 & 1 & 0 \end{pmatrix} \tag{2.29}$$
$$\begin{pmatrix} 0 & 0 & x \\ 0 & 0 & y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}\begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \tag{2.30}$$

De modo que (2.27) se puede escribir así:

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & a \\ 0 & 0 & b \end{pmatrix} \tag{2.31}$$
$$= \begin{pmatrix} 1 \\ 0 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix}\begin{pmatrix} 0 & 1 & 0 \end{pmatrix} + \begin{pmatrix} a \\ b \end{pmatrix}\begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \tag{2.32}$$

vemos que los vectores $\left(\mathbf{v}^i\right)^\top$ (en la notación de (2.19)) expanden todo $\mathbb{R}^{1\times 3}$, de hecho constituyen la base canónica. Pero los vectores $\mathbf{w}_i$ (en la notación de (2.19)) son linealmente dependientes. Podemos usar las propiedades (1.12) del producto outer hasta llegar a una expresión que sea una suma de productos outer en la que tanto los vectores columna como los fila sean linealmente independientes entre sí. Por ejemplo, el tercer sumando en (2.32) se puede expresar así:

$$\begin{pmatrix} a \\ b \end{pmatrix}\begin{pmatrix} 0 & 0 & 1 \end{pmatrix} = a\begin{pmatrix} 1 \\ 0 \end{pmatrix}\begin{pmatrix} 0 & 0 & 1 \end{pmatrix} + b\begin{pmatrix} 0 \\ 1 \end{pmatrix}\begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \tag{2.33}$$

entonces

$$A = \begin{pmatrix} 1 \\ 0 \end{pmatrix}\left[\begin{pmatrix} 1 & 0 & 0 \end{pmatrix} + a\begin{pmatrix} 0 & 0 & 1 \end{pmatrix}\right] + \begin{pmatrix} 0 \\ 1 \end{pmatrix}\left[\begin{pmatrix} 0 & 1 & 0 \end{pmatrix} + b\begin{pmatrix} 0 & 0 & 1 \end{pmatrix}\right] \tag{2.34}$$
$$= \begin{pmatrix} 1 \\ 0 \end{pmatrix}\begin{pmatrix} 1 & 0 & a \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix}\begin{pmatrix} 0 & 1 & b \end{pmatrix} \tag{2.35}$$

en esta última expresión, tanto los vectores $\mathbf{w}_{1,i} \in \mathbb{R}^{2\times 1}$ como los vectores $\left(\mathbf{v}^{1,i}\right)^\top \in \mathbb{R}^{1\times 3}$ son linealmente independientes. Ahora usamos la notación (2.19) completa (diferenciando entre $\mathbf{v}^{1,\cdot}$ y $\mathbf{v}^{2,\cdot}$), porque llegamos a una expresión en la que tanto los vectores $\mathbf{w}_{1,i}$ como los $\mathbf{v}^{1,i}$ son linealmente independientes entre sí, como presupone la expresión (2.19). Vemos que estos últimos expanden un subespacio 2-D en $\mathbb{R}^{3\times 1}$, y que esta dimension es la misma que la del subespacio columna spaneado por los vectores $\mathbf{w}_{1,i}$.

Entonces los vectores $\mathbf{v}^{1,1}$ y $\mathbf{v}^{1,2}$ son los transpuestos de los vectores fila que aparecen en (2.35):

$$\mathbf{v}^{1,1} = \begin{pmatrix} 1 \\ 0 \\ a \end{pmatrix}, \qquad \mathbf{v}^{1,2} = \begin{pmatrix} 0 \\ 1 \\ b \end{pmatrix} \tag{2.36}$$

el span de estos dos vectores es V_1 (ver figura 2). Sabemos que V_2 es ortogonal a V_1, y también sabemos que si $V = V_1 \oplus V_2$, los subespacios V_1 y V_2 son cerrados respecto de la operación de dualidad de una base dada (ver (2.8-2.10)). Como $V = \mathbb{R}^3$ y V_1 es 2-D, V_2 tiene que ser 1-D, por lo que $\mathbf{v}^{2,3}$ y $\mathbf{v}_{2,3}$ tienen que ser colineales y tales que su producto escalar es 1. Los podemos elegir de norma 1 de modo que $\mathbf{v}^{2,3} = \mathbf{v}_{2,3}$ en el subespacio ortogonal al plano spaneado por $\mathbf{v}^{1,1}$ y $\mathbf{v}^{1,2}$. Entonces:

$$\mathbf{v}^{2,3} = \mathbf{v}_{2,3} = \frac{a}{\sqrt{1 + a^2 + b^2}} \begin{pmatrix} -1 \\ -b/a \\ 1/a \end{pmatrix} \tag{2.37}$$

$\mathbf{v}^{1,1}$ y $\mathbf{v}^{1,2}$ en (2.36), y $\mathbf{v}^{2,3}$ en (2.37), forman la base dual asociada a (2.21), de donde se puede obtener fácilmente la base $\{\mathbf{v}_{1,1}, \mathbf{v}_{1,2}, \mathbf{v}_{2,3}\}$. Por otro lado, de (2.35) vemos que la base $\{\mathbf{w}_1, \mathbf{w}_2\}$ es la base canónica, con lo que resolvimos el problema.

En la figura 4 está la solución en Python con el uso de la función **scipy.linalg.null_space()**. Esta función, como su nombre lo sugiere, calcula el espacio nulo de una matriz (array) A devolviendo una matriz B tal que $AB = 0$. De (1.19) vemos que eso significa que cada columna de B es ortogonal a cada fila de A. Entonces, una vez conocida la base $\{\mathbf{v}^{1,1}, \mathbf{v}^{1,2}, \mathbf{v}^{2,3}\}$ en (2.36) y (2.37), aplicando sucesivamente **scipy.linalg.null_space()** a: **1)** la matriz cuyas filas son $\{\mathbf{v}^{1,2}, \mathbf{v}^{2,3}\}$, obteniendo un vector $\tilde{\mathbf{v}}_{1,1}$ en la dirección de $\mathbf{v}_{1,1}$, **2)** la matriz cuyas filas son $\{\mathbf{v}^{1,1}, \mathbf{v}^{2,3}\}$, obteniendo un vector $\tilde{\mathbf{v}}_{1,2}$ en la dirección de $\mathbf{v}_{1,2}$, y **3)** la matriz cuyas filas son $\{\mathbf{v}^{1,1}, \mathbf{v}^{1,2}\}$, obteniendo un vector $\tilde{\mathbf{v}}_{2,3}$ en la dirección de $\mathbf{v}_{2,3}$.

Finalmente:

$$\mathbf{v}_{a,i} = \frac{\tilde{\mathbf{v}}_{a,i}}{\tilde{\mathbf{v}}_{a,i}^\top \mathbf{v}^{a,i}}, \Rightarrow \mathbf{v}_{a,i}^\top \mathbf{v}^{a,i} = 1 \tag{2.38}$$

2.4.4 Dada una matriz A, encontrar la función lineal f_A correspondiente: la solución

Destilemos los pasos del algoritmo general subyacente a la solución del problema que acabamos de hacer para el ejemplo (2.27). Este algoritmo general es el que torna obvio que el rango columna es igual al rango fila en cualquier matriz.

Primer paso

```python
a=1
b=1

v11d = np.array([[1],[0],[a]])
v12d = np.array([[0],[1],[b]])
v23d = np.array([[1],[b/a],[-1/a]])*a/np.sqrt(1+a**2+b**2)

w1 = np.array([[1],[0]])
w2 = np.array([[0],[1]])

A = np.dot(w1,np.transpose(v11d)) + np.dot(w2,np.transpose(v12d))
print("A =",A)

v11 = scipy.linalg.null_space(np.transpose(np.concatenate((v12d, v23d), axis=1)))
v11 /= np.dot(np.transpose(v11),v11d)
v12 = scipy.linalg.null_space(np.transpose(np.concatenate((v11d, v23d), axis=1)))
v12 /= np.dot(np.transpose(v12),v12d)
v23 = scipy.linalg.null_space(np.transpose(np.concatenate((v11d, v12d), axis=1)))
# v23 ya esta automaticamente normalizado

print("A v11 =", np.dot(A,v11))
print("A v12 =", np.dot(A,v12))
print("A v23 =", np.dot(A,v23))
```

```
A = [[1 0 1]
 [0 1 1]]
A v11 = [[ 1.00000000e+00]
 [-1.66533454e-16]]
A v12 = [[5.55111512e-17]
 [1.00000000e+00]]
A v23 = [[0.]
 [0.]]
```

Figure 4: Generación de la base $\{\mathbf{v}_{1,1}, \mathbf{v}_{1,2}, \mathbf{v}_{2,1}\}$ y comprobación de que $A\mathbf{v}_{1,i} = \mathbf{w}_{1,i}$ y $A\mathbf{v}_{2,1} = \mathbf{0}$.

Realizado en (2.31-2.32) en el ejemplo. En general, consiste en reconocer que toda matriz de $\mathbb{R}^{m\times n}$ se puede escribir como suma de productos outer. Si $n > m$, escribimos la matriz así:

$$A = \sum_{i=1}^{n} \mathbf{c}_i\,(\hat{\mathbf{e}}_i)^\top \tag{2.39}$$

donde $\mathbf{c}_i$ es la i-ésima columna de A y $\hat{\mathbf{e}}_i$ es el vector i-ésimo de la base canónica de $\mathbb{R}^n$ (vector columna de con $n-1$ ceros y un 1 en la componente i).

Notar que todos los $\hat{\mathbf{e}}_i$ que aparecen en (2.39) son linealmente independientes entre sí y constituyen la base canónica completa de $\mathbb{R}^n$, tal como ocurre en (2.32) con $\mathbb{R}^3$ en el ejemplo analizado. Sin embargo, los vectores $\mathbf{c}_i \in \mathbb{R}^m$, como por hipótesis $m < n$ y hay n de ellos, son necesariamente linealmente dependientes.

Si $m > n$ la escribimos así:

$$A = \sum_{i=1}^{m} \hat{\mathbf{e}}_i\,(\mathbf{f}_i)^\top \tag{2.40}$$

donde $\hat{\mathbf{e}}_i$ es el vector i-ésimo de la base canónica de $\mathbb{R}^m$ y $(\mathbf{f}_i)^\top$ la i-ésima fila ($\in \mathbb{R}^{1\times n}$) de A.

Los $\hat{\mathbf{e}}_i$ en (2.40) son linealmente independientes entre sí y constituyen la base canónica completa de $\mathbb{R}^m$, mientras que los vectores $\mathbf{f}_i \in \mathbb{R}^n$, como por hipótesis $m > n$ y hay m de ellos, son necesariamente linealmente dependientes.

Si $m = n$, podemos usar (2.39) o (2.39) indistintamente.

En el ejemplo (2.27), como $n = 3 > 2 = m$, usamos (2.39).

Esto finaliza el primer paso, en el que explotamos el hecho de que **toda matriz** $A \in \mathbb{R}^{m\times n}$ **se puede escribir como una suma de productos outer**. Se elige la suma, (2.39) si $n > m$. Las filas que aparecen en dicha suma son linealmente independientes. Se elige la suma, (2.40) si $m > n$. Las columnas que aparecen en dicha suma son linealmente independientes. Se puede elegir tanto (2.39) como (2.40) si $m = n$. Si se elige (2.39) las filas son linealmente independientes, y si se elige (2.40) las columnas son linealmente independientes.

Segundo paso

Realizado en ejemplo en las ecuaciones (2.33-2.35). Usamos las propiedades (1.12) del producto outer (particularmente las tres últimas), para eliminar todas las columnas y/o filas linealmente dependientes, de modo que el resultado es una suma de productos outer en la que todas las columnas son linealmente independientes entre sí y lo mismo todas las filas. Recordemos las tres últimas propiedades (1.12) en notación matricial:

$$(\mathbf{y}+\mathbf{z})\mathbf{x}^\top = \mathbf{y}\mathbf{x}^\top + \mathbf{z}\mathbf{x}^\top \tag{2.41}$$

$$\mathbf{x}(\mathbf{y}+\mathbf{z})^\top = \mathbf{x}\mathbf{y}^\top + \mathbf{x}\mathbf{z}^\top \tag{2.42}$$

$$c(\mathbf{x}\mathbf{y}^\top) = (c\mathbf{x})\mathbf{y}^\top = \mathbf{x}(c\mathbf{y}^\top) \tag{2.43}$$

Con (2.42) y 2.43) vemos que si $\{\mathbf{x}_1, \mathbf{x}_1, \mathbf{x}_3\}$ son linealmente independientes como en (2.39), e $\mathbf{y}_3$ es linealmente dependiente de $\mathbf{y}_1$ y $\mathbf{y}_2$, esto es, si $\mathbf{y}_3 = b_1\mathbf{y}_1 + b_2\mathbf{y}_2$, para $b_1, b_2 \in \mathbb{R}$, no ambas simultáneamente cero,

$$\mathbf{y}_1\mathbf{x}_1^\top + \mathbf{y}_2\mathbf{x}_2^\top + \mathbf{y}_3\mathbf{x}_3^\top = \mathbf{y}_1(\mathbf{x}_1^\top + b_1\mathbf{x}_3^\top) + \mathbf{y}_2(\mathbf{x}_2^\top + b_2\mathbf{x}_3^\top) \tag{2.44}$$

Esta ecuación nos dice que a pesar que del lado izquierdo tenemos tres filas linealmente independientes, si una de las columnas es linealmente dependiente de las otras dos, como ocurre en (2.39), las podemos re-arreglar de modo que queden dos columnas y dos filas. Dado que por hipótesis, los vectores $\mathbf{x}_1$, $\mathbf{x}_2$ e $\mathbf{x}_3$ son linealmente independientes, entonces $\mathbf{x}_1 + b_1\mathbf{x}_3$ y $\mathbf{x}_2 + b_2\mathbf{x}_3$ también lo son.

Si quedan más columnas linealmente dependientes entre sí, continuamos con este procedimiento hasta que todas sean linealmente independientes. Notar que en cada iteración de este procedimiento, los vectores fila $\mathbf{x}_i^\top$ resultantes son necesariamente linealmente independientes ya que en la iteración anterior lo eran.

En cada iteración tenemos menos términos en la suma, pero como cada término tiene una columna y una fila, al final del proceso terminamos con igual número de columnas que de filas linealmente independientes.

De manera análoga, con (2.41) y 2.43) vemos que si $\{\mathbf{y}_1, \mathbf{y}_1, \mathbf{y}_3\}$ son linealmente independientes, y $\mathbf{x}_3$ es linealmente dependiente de $\mathbf{x}_1$ y $\mathbf{x}_2$, esto es, si $\mathbf{x}_3 = a_1\mathbf{x}_1 + a_2\mathbf{x}_2$, para $a_1, a_2 \in \mathbb{R}$, no ambas simultáneamente cero,

$$\mathbf{y}_1\mathbf{x}_1^\top + \mathbf{y}_2\mathbf{x}_2^\top + \mathbf{y}_3\mathbf{x}_3^\top = (\mathbf{y}_1 + a_1\mathbf{y}_3)\mathbf{x}_1^\top + (\mathbf{y}_2 + a_2\mathbf{y}_3)\mathbf{x}_2^\top \tag{2.45}$$

Como los vectores $\{\mathbf{y}_1, \mathbf{y}_1, \mathbf{y}_3\}$ son linealmente independientes, las dos columnas que aparecen en el lado derecho de (2.45), $\mathbf{y}_1 + a_1\mathbf{y}_3$ y $\mathbf{y}_2 + a_2\mathbf{y}_3$, también lo son.

Si quedan más filas linealmente dependientes entre sí, de manera anaaloga a como hicimos antes, continuamos con este procedimiento hasta que todas sean linealmente independientes.

Al finalizar este procedimiento, la matriz A queda expresada como una suma de productos

$$A = \sum_{i=1}^{j} \mathbf{y}_i\mathbf{x}_i^\top \tag{2.46}$$

donde los vectores $\{\mathbf{y}_i\} \in \mathbb{R}^m$ son linealmente independientes entre sí, y los vectores $\{\mathbf{x}_i\} \in \mathbb{R}^n$ también lo son.

Tercer paso

Identificamos en (2.46) a:

$$\mathbf{y}_i = \mathbf{w}_{1,i}, \quad i = 1, \dots j \tag{2.47}$$

$$\mathbf{x}_i = \mathbf{v}^{1,i}, \quad i = 1, \dots j \tag{2.48}$$

por lo que ya identificamos a los vectores $\{\mathbf{w}_{1,1}, \dots, \mathbf{w}_{1,j}\}$, base de W_1 en (2.22), ver figura 2. Los vectores $\{\mathbf{w}_{2,j+1}, \dots, \mathbf{w}_{2,m}\}$ son cualquier base del complemento ortogonal de W_1, $W_1^\perp$, que se obtiene en la práctica con **scipy.linalg.null_space()** aplicada a la matriz cuyas filas son los $\mathbf{w}_{1,i}^\top$, $i = 1, \dots j$. Con esto identificamos completamente a la base W en (2.22).

Construyendo una matriz cuyas filas son $\left(\mathbf{v}^{1,i}\right)^\top$, con los $\mathbf{v}^{1,i}$ identificados en (2.48), y calculando su espacio nulo (en la práctica, por ejemplo, aplicando **scipy.linalg.null_space()** a dicha matriz), obtenemos la base $\{\mathbf{v}_{2,j+1}, \dots, \mathbf{v}_{2,n}\}$ de V_2 correspondiente a la segunda parte de (2.21), ver figura 2.

Notar que cualquier base de V_2 puede actuar como $\{\mathbf{v}_{2,j+1}, \dots, \mathbf{v}_{2,n}\}$, lo importante es que spanee a todo V_2, ya que necesitamos que los $\mathbf{v}_{1,i}$ sean ortogonales a dicho subespacio.

Calculamos, para $i = 1, \dots, j$, un representante de los espacios nulos de la matriz cuyas filas son

$$\tilde{\mathbf{v}}_{1,i} = \text{ rep } \mathcal{N}\left\{\left(\mathbf{v}^{1,1}\right)^\top, \dots, \left(\mathbf{v}^{1,i-1}\right)^\top, \left(\mathbf{v}^{1,i+1}\right)^\top, \dots, \left(\mathbf{v}^{1,j}\right)^\top, \mathbf{v}_{2,j+1}^\top, \dots, \mathbf{v}_{2,n}^\top\right\} \tag{2.49}$$

donde a los $\mathbf{v}^{1,s}$, $s = 1, \dots, j$ los extraemos de (2.48), y acabamos de explicar cómo obtener los $\mathbf{v}^{2,k}$, $k = j+1, \dots, n$. Noten que para cada i, el espacio nulo de la matriz indicada en (2.49) (que se puede calcular en la práctica aplicando **scipy.linalg.null_space()** a dicha matriz) tiene una y solo una dimensión, por lo que $\tilde{\mathbf{v}}_{1,i}$ es colineal con $\mathbf{v}_{1,i}$.

Finalmente:

$$\mathbf{v}_{1,i} = \frac{\tilde{\mathbf{v}}_{1,i}}{\tilde{\mathbf{v}}_{1,i}^{\top}\mathbf{v}^{1,i}}, \quad i = 1, \ldots, j \tag{2.50}$$

de modo que $\mathbf{v}_{1,i}^{\top}\mathbf{v}^{1,i} = 1,$.

Con esto quedan completamente identificadas bases (2.21-2.22) tal que la matriz A de la que partimos actúa de acuerdo a (2.23-2.24). De modo que queda definida completamente la función $\mathbf{f}_A\colon \mathbb{R}^{n\times 1} \to \mathbb{R}^{m\times 1}$ que mapea a los vectores de (2.21) de acuerdo a (2.25-2.26). ✓

2.5 Igualdad entre rango fila y rango columna de una matriz

Recordemos que el objetivo de esta sección es encontrar la manera de mirar las cosas de modo que el hecho conocido de que para toda matriz, el rango fila es igual al rango columna, resulte obvio. El camino planteado para llegar al objetivo, fue elegido para cumplir además con objetivos secundarios. Entre ellos, mostrar la relación entre matrices concretas y funciones lineales abstractas en sección 2.2, bases duales en la sección 2.3, la expresión de matrices como suma de productos outer, etc.

El "segundo paso" en 2.4.4, que finaliza con la expresión (2.46) para $A \in R^{m\times n}$ como suma de productos outer, donde tanto los j vectores $\{\mathbf{y}_i\} \in \mathbb{R}^m$ como los j vectores $\{\mathbf{x}_i\} \in \mathbb{R}^n$ son linealmente independientes, y probablemente nunca vayamos a implementar en la práctica a menos que alguien lo solicite explícitamente (ver sección 2.8), es crucial para que resulte obvio entender que el rango fila y el rango columna de cualquier matriz son iguales.

Sin embargo, la igualdad del número de filas y columnas linealmente independientes en (2.46) es "casi" el resultado buscado. Pero es importante notar que las columnas que aparecen en dicha suma no son necesariamente idénticas a las columnas de A. Por ejemplo, la tercera columna en la matriz (2.27) no aparece en la expresión (2.35), que es de la forma (2.46).

De la misma manera, las filas que aparecen en (2.46) no son necesariamente iguales a las filas de A.

La ecuación (2.46) es una expresión con igual número de filas que de columnas linealmente independientes. Pero falta probar que el subespacio espaneado por las columnas de A es idéntico al spaneado por las columnas que aparecen en (2.46). Y que el subespacio espaneado por las filas de A es idéntico al spaneado por las filas que aparecen en (2.46). Probamos estas dos propiedades a continuación.

El subespacio espaneado por las columnas de A es idéntico al spaneado por las columnas que aparecen en (2.46):

La manera (1.14) de mirar el producto matriz-vector indica que el resultado de multiplicar $A\mathbf{v}, \forall \mathbf{v} \in V$ es el span de las columnas de A. (2.46) indica que el resultado es el span de los vectores columna que aparecen en dicha ecuación. Son dos maneras distintas de calcular lo mismo, por lo que tenemos la prueba deseada. ✓

El subespacio espaneado por las filas de A es idéntico al spaneado por las filas que aparecen

en (2.46):

(2.46) indica que $A\mathbf{v} = \mathbf{0}$ sólo para vectores ortogonales a las filas que aparecen en dicha ecuación. La manera (1.13) de visualizar el producto matriz-vector nos indica que $A\mathbf{v} = \mathbf{0}$ sólo para vectores ortogonales a las filas de A. Por lo que el espacio nulo de las filas que aparecen en (2.46) es idéntico al de las filas de A. El complemento ortogonal de este espacio nulo es entonces el span de las filas que aparecen en (2.46) y el span de las filas de A, que por lo tanto son idénticos como queríamos probar. ✓

Como la dimension de los subespacio spaneado por las filas y columnas que aparecen en (2.46) son iguales, obtenemos el resultado deseado. ✓

2.6 Propiedades del rango una matriz

- Para $A \in \mathbb{R}^{m\times n}$, $\mathrm{rank}(A) \leq \min(m, n)$. Si $\mathrm{rank}(A) = \min(m, n)$, entonces se dice que A es *full rank*.
- Para $A \in \mathbb{R}^{m\times n}$, $\mathrm{rank}(A) = \mathrm{rank}\left(A^\top\right)$
- Para $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{n\times p}$, $\mathrm{rank}(AB) \leq \min(\mathrm{rank}(A), \mathrm{rank}(B))$
- Para $A, B \in \mathbb{R}^{m\times n}$, $\mathrm{rank}(A + B) \leq \mathrm{rank}(A) + \mathrm{rank}(B)$

Ejercicio: A partir de los conceptos aprendidos, “probar con palabras” estas propiedades.

2.7 Observaciones generales sobre funciones lineales, matrices, y bases duales

Lo explorado en esta sección nos brinda una perspectiva profunda de lo que en el fondo son las funciones lineales. El hecho de que $\dim V_1 = \dim W_1$ nos indica que en cierto sentido las funciones lineales son biyecciones. En un sentido abstracto, dada una función lineal, podríamos descartar a V_2 y a W_2, que son simplemente artefactos del espacio en el que estamos *realizando* la función. Entonces, en un sentido abstracto, toda función lineal es una biyección entre espacios vectoriales (V_1 y W_1) de igual dimensión.

Llevado a la realización de funciones lineales en términos de matrices, el párrafo anterior significa que, en el fondo, toda matriz esconde una matriz cuadrada de rango completo.

Por otro lado, hay situaciones en las que el bakground donde las funciones (o matrices) actúan es importante. Es decir, situaciones en las que V_2 y/o W_2 contienen parte de la información que necesitamos para entender el problema. En esos casos es fundamental entender que toda función lineal (o matriz) $\mathbf{f}\colon V \to W$ separa naturalmente a V en $V_2 = \mathcal{N}(\mathbf{f})$ y $V_1 = V_2^\perp$, de modo que $V = V_1 \oplus V_2$. Y, dados el codominio W y el rango W_1, definimos $W_2 = W_1^\perp$, de modo que $W = W_1 \oplus W_2$.

La base dual de una base dada es un concepto tremendamente poderoso para obtener resultados de manera que resulten “obvios”. Además, el mecanismo de subir y bajar indices, es como que toma vida propia y funciona en automático, como vamos a ver más adelante.

Como la base dual de la base dual de una base dada es la base original, pierde sentido (por lo menos en álgebra lineal) la noción de a cuál llamamos dual y a cuál no. Lo importante es que, dada una base linealmente independiente de un espacio vectorial V, a la que denotamos con subíndices,

$$\{\mathbf{v}_1, \ldots, \mathbf{v}_n\} : \text{ base de } V \text{ denotada con subíndices} \tag{2.51}$$

siempre existe otra base de V linealmente independiente

$$\left\{\mathbf{v}^1, \ldots, \mathbf{v}^n\right\} : \text{ base de } V \text{ denotada con superíndices} \tag{2.52}$$

tal que

$$\mathbf{v}_i^\top \mathbf{v}^j = \left(\mathbf{v}^j\right)^\top \mathbf{v}_i = \delta_i^j \tag{2.53}$$

notemos que la “delta” en esta ecuación tiene el índice j arriba y el i abajo, tal como aparecen en el lado izquierdo de la ecuación. Esto lo hacemos para llevar el registro de los índices, pero esa delta significa lo mismo que la usada previamente con dos subíndices: cero si $i \neq j$ y uno si $i = j$.

Vamos a adoptar la siguiente convención: dado cualquier vector $\mathbf{x}$, a sus coordenadas en la base con subíndices las expresamos con superíndices, y a sus coordenadas en la base de superíndices la expresamos con subíndices:

$$\mathbf{x} = \sum_{i=1}^n x^i \mathbf{v}_i = \sum_{i=1}^n x_i \mathbf{v}^i \tag{2.54}$$

Notar que

$$\mathbf{x}^\top \mathbf{x} = \left(\sum_{i=1}^n x^i \mathbf{v}_i^\top\right)\left(\sum_{j=1}^n x_i \mathbf{v}^i\right) = \sum_{i=1}^n \sum_{j=1}^n x^i x_j \left(\mathbf{v}_i^\top \mathbf{v}^j\right) = \sum_{i=1}^n x^i x_i \tag{2.55}$$

donde en la última igualdad usamos (2.53). Este resultado es reminiscente del cuadrado de la norma ℓ_2 de un vector en una base ortonormal, es decir, del teorema de Pitágoras. Justamente, una de las cosas que permite trabajar con esas dos bases es que extiende la simplicidad formal de las bases ortonormales a bases no ortonormales!

La ecuación (2.55) aplica también a productos escalares entre dos vectores diferentes: $\mathbf{x}^\top \mathbf{y} = \sum_{i=1}^n x^i y_i = \sum_{i=1}^n x_i y^i$. Nuevamente un resultado que extiende la simplicidad formal de las bases ortonormales a bases no ortonormales.

Por último, recordamos que dado cualquier espacio vectorial V, un subespacio V_1 y su complemento ortogonal $V_2 = V_1^\perp$, si $\{\mathbf{v}_{1,1}, \ldots, \mathbf{v}_{1,j}, \mathbf{v}_{2,j+1}, \ldots, \mathbf{v}_{2,n}\}$ es una base de V tal que $\{\mathbf{v}_{1,1}, \ldots, \mathbf{v}_{1,j}\}$ es base de V_1 y $\{\mathbf{v}_{2,j+1}, \ldots, \mathbf{v}_{2,n}\}$ es base de V_2, es decir, si

$$\mathbf{v}_{1,i}^\top \mathbf{v}_{2,k} = 0, \forall i = 1, \ldots, j, \text{ and } k = j+1, \ldots, n \tag{2.56}$$

entonces la base $\{\mathbf{v}^{1,1}, \ldots, \mathbf{v}^{1,j}, \mathbf{v}^{2,j+1}, \ldots, \mathbf{v}^{2,n}\}$ es una base de V tal que $\{\mathbf{v}^{1,1}, \ldots, \mathbf{v}^{1,j}\}$ es base de V_1 y $\{\mathbf{v}^{2,j+1}, \ldots, \mathbf{v}^{2,n}\}$ es base de $V_2 = V_1^\perp$, es decir,

$$\left(\mathbf{v}^{1,i}\right)^\top \mathbf{v}^{2,k} = \left(\mathbf{v}^{1,i}\right)^\top \mathbf{v}_{2,k} = \left(\mathbf{v}_{1,i}\right)^\top \mathbf{v}^{2,k} = 0, \forall i = 1, \ldots, j, \text{ and } k = j+1, \ldots, n \tag{2.57}$$

2.8 Ejercicios

1) Generar al azar en un Jupyter notebook 4 vectores $\in \mathbb{R}^{4\times 1}$, $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ y 2 vectores $\in \mathbb{R}^{3\times 1}$, $\{\mathbf{w}_1, \mathbf{w}_2\}$. Construir la matriz de 3×4 que genera el siguiente mapping:

$$\mathbf{f}(\mathbf{v}_1) = \mathbf{w}_1 \tag{2.58}$$
$$\mathbf{f}(\mathbf{v}_2) = \mathbf{w}_2 \tag{2.59}$$
$$\mathbf{f}(\mathbf{v}_3) = \mathbf{0} \tag{2.60}$$
$$\mathbf{f}(\mathbf{v}_4) = \mathbf{0} \tag{2.61}$$

2) Generar al azar en un Jupyter notebook 4 una matriz $\in \mathbb{R}^{2\times 4}$. Con probabilidad ~ 1 las dos filas $\in \mathbb{R}^{1\times 4}$ van a ser linealmente independientes (por qué?) y de las 4 columnas $\in \mathbb{R}^{2\times 1}$, 2 van a ser linealmente dependientes de las otras dos. Encontrar dos bases, $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ y 2 vectores $\in \mathbb{R}^{3\times 1}$, y $\{\mathbf{w}_1, \mathbf{w}_2\}$, que definan unívocamente a la función en el sentido de las ecuaciones (2.21-2.26).

3 Más sobre bases duales y aplicaciones

Esta sección, que como su nombre lo indica, esta dedicada a bases duales y sus aplicaciones, pretende ser semi-independiente de la anterior, para que sea útil para lectores que quieran concentrarse exclusivamente en aplicaciones. Por lo que algunos resultados aquí presentados repiten parcialmente resultados presentados en la sección anterior.

3.1 Coordenadas de vectores en bases duales

Supongamos que tenemos bases linealmente independientes y duales una de la otra, de un espacio vectorial V

$$\left\{\mathbf{x}_{1,1}, \ldots, \mathbf{x}_{1,j}, \mathbf{x}_{2,j+1}, \ldots, \mathbf{x}_{2,n}\right\} \tag{3.1}$$
$$\left\{\mathbf{x}^{1,1}, \ldots, \mathbf{x}^{1,j}, \mathbf{x}^{2,j+1}, \ldots, \mathbf{x}^{2,n}\right\}, \quad 1 \le j \le n \tag{3.2}$$

donde $\left\{\mathbf{x}_{1,1}, \ldots, \mathbf{x}_{1,j}\right\}$ es base de un subespacio V_1 y $\left\{\mathbf{x}_{2,j+1}, \ldots, \mathbf{x}_{2,n}\right\}$ es base del subespacio V_2. Cada uno de esos subespacios es complemento ortogonal del otro en V: $V_2 = V_1^\perp$, $V_1 = V_2^\perp$, o $V = V_1 \oplus V_1$. Es decir que tenemos las siguientes productos escalares:

$$(\mathbf{x}_{1,i})^\top \mathbf{x}_{2,k} = \left(\mathbf{x}^{1,i}\right)^\top \mathbf{x}^{2,k} = 0, \quad \forall i = 1, \ldots, j, \quad k = j+1, \ldots, n, \quad \text{porque} \quad V_2 = V_1^\perp \tag{3.3}$$
$$(\mathbf{x}_{a,i})^\top \mathbf{x}^{b,k} = \left(\mathbf{x}^{a,i}\right)^\top \mathbf{x}_{b,k} = \delta_{ab}\delta_{ik}, \quad \forall i,k = 1, \ldots, n, \quad \text{porque son bases duales} \tag{3.4}$$

Proposición: $\left\{\mathbf{x}^{1,1}, \ldots, \mathbf{x}^{1,j}\right\}$ es base de V_1 y $\left\{\mathbf{x}^{2,j+1}, \ldots, \mathbf{x}^{2,n}\right\}$ es base de V_2.

Para probar la primera afirmación basta ver que por (3.4), cualquier $\mathbf{x}^{1,i}$, $i = 1, \ldots, j$, es ortogonal a la base de V_2 $\left\{\mathbf{x}_{2,j+1}, \ldots, \mathbf{x}_{2,n}\right\}$. Y para probar la segunda, por la misma razón, todo $\mathbf{x}^{1,k}$, $k = j+1, \ldots, n$, es ortogonal a la base de V_1 $\left\{\mathbf{x}_{1,1}, \ldots, \mathbf{x}_{1,j}\right\}$.✓

Como ambos conjuntos, (3.1) y (3.2), son bases de V linealmente independientes, cualquier vector $\mathbf{y} \in V$ se puede expandir de manera única en términos de ambas. Llamemos $y^{a,i}$ a las coordenadas de $\mathbf{y}$ con respecto a la base (3.1), e $y_{b,k}$ a sus coordenadas con respecto a la base (3.2):

$$\mathbf{y} = \sum_{i=1}^{j} y^{1,i}\mathbf{x}_{1,i} + \sum_{k=j+1}^{n} y^{2,k}\mathbf{x}_{2,k} \tag{3.5}$$

$$= \sum_{i=1}^{j} y_{1,i}\mathbf{x}^{1,i} + \sum_{k=j+1}^{n} y_{2,k}\mathbf{x}^{2,k} \tag{3.6}$$

notar que los índices de las coordenadas son superíndices si los de la correspondiente base son subíndices y viceversa.

Multiplicando escalarmente a $\mathbf{y}$ por elementos de ambas bases, (3.3) y (3.4) implican las siguientes relaciones entre coordenadas:

$$y^{1,r} = \left(\mathbf{x}^{1,r}\right)^\top \mathbf{y} = \sum_{i=1}^{j} y_{1,i}\left(\left(\mathbf{x}^{1,r}\right)^\top \mathbf{x}^{1,i}\right) \tag{3.7}$$

$$y^{2,s} = \left(\mathbf{x}^{2,s}\right)^\top \mathbf{y} = \sum_{k=j+1}^{n} y_{2,k}\left(\left(\mathbf{x}^{2,s}\right)^\top \mathbf{x}^{2,k}\right) \tag{3.8}$$

$$y_{1,r} = \left(\mathbf{x}_{1,r}\right)^\top \mathbf{y} = \sum_{i=1}^{j} y^{1,i}\left(\left(\mathbf{x}_{1,r}\right)^\top \mathbf{x}_{1,i}\right) \tag{3.9}$$

$$y_{2,s} = \left(\mathbf{x}_{2,s}\right)^\top \mathbf{y} = \sum_{k=j+1}^{n} y^{2,k}\left(\left(\mathbf{x}_{2,s}\right)^\top \mathbf{x}_{2,k}\right) \tag{3.10}$$

3.2 La identidad y proyectores sobre subespacios como suma de productos outer

Pensemos en $\mathbb{R}^2$ y la base canónica $\{\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2\}$:

$$\hat{\mathbf{e}}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad \hat{\mathbf{e}}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{3.11}$$

Consideremos las siguientes matrices:

$$\hat{\mathbf{e}}_1\hat{\mathbf{e}}_1^\top = \begin{pmatrix} 1 \\ 0 \end{pmatrix}(1,0) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad \hat{\mathbf{e}}_2\hat{\mathbf{e}}_2^\top = \begin{pmatrix} 0 \\ 1 \end{pmatrix}(0,1) = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \tag{3.12}$$

vemos que $\hat{\mathbf{e}}_1\hat{\mathbf{e}}_1^\top$ "proyecta" sobre la dirección $\hat{\mathbf{e}}_1$, en el sentido que si tengo un vector $\mathbf{y} = (y_1, y_2)^\top$, su proyección en la dirección $\hat{\mathbf{e}}_1$ es obviamente $(y_1, 0)^\top$, y eso es lo que logra $\hat{\mathbf{e}}_1\hat{\mathbf{e}}_1^\top$ cuando multiplica al vector $\mathbf{y}$. De la misma manera, $\hat{\mathbf{e}}_2\hat{\mathbf{e}}_2^\top$ proyecta sobre la dirección $\hat{\mathbf{e}}_2$:

$$\hat{\mathbf{e}}_1\hat{\mathbf{e}}_1^\top \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} y_1 \\ 0 \end{pmatrix}, \qquad \hat{\mathbf{e}}_2\hat{\mathbf{e}}_2^\top \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 0 \\ y_2 \end{pmatrix} \tag{3.13}$$

además, de (3.12) vemos que podemos escribir la identidad como:

$$I = \sum_{i=1}^{2} \hat{\mathbf{e}}_i\hat{\mathbf{e}}_i^\top = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{3.14}$$

Todas estas ecuaciones extienden de manera trivial a $\mathbb{R}^n$.

Tratemos de ver esto de una manera un poco más abstracta. El vector $\mathbf{y} = (y_1, y_2)^\top$ se puede escribir así:

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = y_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + y_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = y_1\hat{\mathbf{e}}_1 + y_2\hat{\mathbf{e}}_2 = \sum_{i=1}^{2} y_i\hat{\mathbf{e}}_i \tag{3.15}$$

las y_i son las *coordenadas* de $\mathbf{y}$ en la base $\hat{\mathbf{e}}_i$. Notar que están escritas en letras sin negrita, indicando que son números reales. Veamos de manera formal los productos realizados en (3.13):

$$\left(\hat{\mathbf{e}}_j\hat{\mathbf{e}}_j^\top\right)\mathbf{y} = \left(\hat{\mathbf{e}}_j\hat{\mathbf{e}}_j^\top\right)\left(\sum_{i=1}^{2} y_i\hat{\mathbf{e}}_i\right) = \sum_{i=1}^{2} y_i \left(\hat{\mathbf{e}}_j\hat{\mathbf{e}}_j^\top\right)\hat{\mathbf{e}}_i = \sum_{i=1}^{2} y_i\hat{\mathbf{e}}_j \underbrace{\left(\hat{\mathbf{e}}_j^\top\hat{\mathbf{e}}_i\right)}_{=\delta_{ji}} = y_j\hat{\mathbf{e}}_j \quad \checkmark \tag{3.16}$$

Entonces $\hat{\mathbf{e}}_j\hat{\mathbf{e}}_j^\top$ es un proyector sobre el vector $\hat{\mathbf{e}}_j$.

Todo esto funciona porque la base $\{\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2\}$ es ortonormal, por lo que el producto escalar $\hat{\mathbf{e}}_j^\top\hat{\mathbf{e}}_i = \delta_{ji}$, lo que genera la última igualdad en (3.16). Si la base considerada no fuera ortonormal, nada de todo esto se puede lograr. Sin embargo, un gran mérito de las bases duales, es que todo esto se puede reproducir de manera casi idéntica con ellas, como probamos a continuación.

Consideremos las bases duales (3.1-3.2). La matriz

$$I = \sum_{i=1}^{j} \mathbf{x}_{1,i}\left(\mathbf{x}^{1,i}\right)^\top + \sum_{k=j+1}^{n} \mathbf{x}_{2,k}\left(\mathbf{x}^{2,k}\right)^\top \tag{3.17}$$

$$= \sum_{i=1}^{j} \mathbf{x}^{1,i}\left(\mathbf{x}_{1,i}\right)^\top + \sum_{k=j+1}^{n} \mathbf{x}^{2,k}\left(\mathbf{x}_{2,k}\right)^\top \tag{3.18}$$

es la matriz identidad,

$$\mathrm{Proj}\left(\cdot; V_1\right) = \sum_{i=1}^{j} \mathbf{x}_{1,i}\left(\mathbf{x}^{1,i}\right)^\top = \sum_{i=1}^{j} \mathbf{x}^{1,i}\left(\mathbf{x}_{1,i}\right)^\top \tag{3.19}$$

es un proyector sobre V_1, y

$$\mathrm{Proj}\,(\cdot; V_2) = \sum_{k=j+1}^{n} \mathbf{x}_{2,k} \left(\mathbf{x}^{2,k}\right)^{\top} = \sum_{k=j+1}^{n} \mathbf{x}^{2,k} \left(\mathbf{x}_{2,k}\right)^{\top} \tag{3.20}$$

es un proyector sobre V_2. Verificamos a continuación (3.17), dejando la verificación de (3.18), (3.19) y (3.20) como ejercicio para el lector.

Recordemos que como las bases (3.1-3.2) son completas y linealmente independientes, cualquier vector $\mathbf{y}$ se puede escribir de forma única en términos de cualquier de ellas, tal como hicimos en (3.5-3.6). Miremos por ejemplo el producto de I, en la forma (3.17), por la expansion (3.5) de $\mathbf{y}$:

$$\begin{aligned}
I\mathbf{y} &= \left(\sum_{i=1}^{j} \mathbf{x}_{1,i} \left(\mathbf{x}^{1,i}\right)^{\top} + \sum_{k=j+1}^{n} \mathbf{x}_{2,k} \left(\mathbf{x}^{2,k}\right)^{\top}\right)\left(\sum_{r=1}^{j} y^{1,r}\mathbf{x}_{1,r} + \sum_{s=j+1}^{n} y^{2,s}\mathbf{x}_{2,s}\right) \qquad (3.21)\\
&= \left(\sum_{i=1}^{j}\sum_{r=1}^{j} y^{1,r}\mathbf{x}_{1,i} \underbrace{\left(\mathbf{x}^{1,i}\right)^{\top}\mathbf{x}_{1,r}}_{=\,\delta_{ir}}\right) + \left(\sum_{i=1}^{j}\sum_{s=j+1}^{n} y^{2,s}\mathbf{x}_{1,i} \underbrace{\left(\mathbf{x}^{1,i}\right)^{\top}\mathbf{x}_{2,s}}_{=\,0}\right) + \\
&\quad \left(\sum_{k=j+1}^{n}\sum_{r=1}^{j} y^{1,r}\mathbf{x}_{2,k} \underbrace{\left(\mathbf{x}^{1,i}\right)^{\top}\mathbf{x}_{1,r}}_{=\,0}\right) + \left(\sum_{k=j+1}^{n}\sum_{s=j+1}^{n} y^{2,s}\mathbf{x}_{2,k} \underbrace{\left(\mathbf{x}^{2,k}\right)^{\top}\mathbf{x}_{2,s}}_{=\,\delta_{ks}}\right)\\
&= \sum_{r=1}^{j} y^{1,r}\mathbf{x}_{1,r} + \sum_{s=j+1}^{n} y^{2,s}\mathbf{x}_{2,s} = \mathbf{y}, \quad \checkmark
\end{aligned}$$

Ejercicio: Consideremos $\mathbb{R}^2$ con la base (3.1) dada por $\{\mathbf{v}_{1,1} = (2,1)^{\top}, \mathbf{v}_{1,2} = (1,2)^{\top}\}$, con $j = n = 2$. 1) Construir la base dual (3.2). 2) Verificar (3.17) y (3.18).

Ejercicio: Realizar una prueba similar para (3.18), (3.19) y (3.20).

3.3 Matrices cuadradas como suma de productos outer

Consideremos una matriz cuadradas $A \in \mathbb{R}^{n\times n}$. Lo visto en la sección 2 implica que existen dos bases linealmente independientes de $\mathbb{R}^{n\times 1}$,

$$\left\{\mathbf{v}_{1,1}, \ldots, \mathbf{v}_{1,j}, \mathbf{v}_{2,j+1}, \ldots, \mathbf{v}_{2,n}\right\} \tag{3.22}$$

$$\left\{\mathbf{w}_{1,1}, \ldots, \mathbf{w}_{1,j}, \mathbf{w}_{2,j+1}, \ldots, \mathbf{w}_{2,n}\right\} \tag{3.23}$$

con $j \leq n$, tal que

$$A = \sum_{i=1}^{j} \mathbf{w}_{1,i} \left(\mathbf{v}^{1,i}\right)^{\top} \tag{3.24}$$

donde la acción de A sobre la base $\{\mathbf{v}\}$ es:

$$A\mathbf{v}_{1,i} = \mathbf{w}_{1,i} \tag{3.25}$$

$$A\mathbf{v}_{2,i} = \mathbf{0} \tag{3.26}$$

$\left\{\mathbf{v}_{2,j+1},\ldots,\mathbf{v}_{2,n}\right\}$ spanea a $V_2 = \mathcal{N}(A)$, y si llamamos V_1 al span de $\left\{\mathbf{v}_{1,1},\ldots,\mathbf{v}_{1,j}\right\}$, $\mathbb{R}^n = V_1 \oplus V_2$.

De manera similar, si llamamos W_1 al span de $\left\{\mathbf{w}_{1,1},\ldots,\mathbf{w}_{1,j}\right\}$, $\dim W_1 = \dim V_1$, y llamando W_2 al span de $\left\{\mathbf{w}_{2,j+1},\ldots,\mathbf{w}_{2,n}\right\}$, elegida de modo que $W_2 = W_1^{\perp}$ (o $W = W_1 \oplus W_2$), como A es cuadrada y ambas bases (3.22) y (3.23) tienen n vectores linealmente independientes, $\dim W_2 = \dim V_2$.

Si $j = n$, la matriz (3.24) tiene "rango completo", o "full rank", y tiene inversa, que es:

$$A^{-1} = \sum_{k=1}^{n} \mathbf{v}_{1,k} \left(\mathbf{w}^{1,k}\right)^{\top} \tag{3.27}$$

multiplicando a (3.24) por derecha o por izquierda por (3.27), teniendo en cuenta (3.4) con $j = n$, llegamos trivialmente a (3.17) o (3.18). Notar que en (3.27), los roles de la base de vectores $\mathbf{w}$ y $\mathbf{v}$ están intercambiados con respecto a (3.24) y además los subíndices pasan a superíndices y vice-versa.

Una dada matriz $n \times n$ de rango completo A, pensada como colección de filas,

$$A = \begin{bmatrix} -\left(\mathbf{a}^1\right)^{\top} - \\ \vdots \\ -\left(\mathbf{a}^m\right)^{\top} - \end{bmatrix} = \sum_{i=1}^{j} \hat{\mathbf{e}}_i \left(\mathbf{a}^i\right)^{\top} \tag{3.28}$$

corresponde a (3.24) con $\mathbf{w}_i = \hat{\mathbf{e}}_i$ y los vectores $\mathbf{v}^i = \mathbf{a}^i$ son simplemente las filas de A. (3.27) indica que su inversa es

$$A^{-1} = \sum_{k=1}^{n} \mathbf{a}_k \left(\hat{\mathbf{e}}^k\right)^{\top} = \begin{bmatrix} | & & | \\ \mathbf{a}_1 & \cdots & \mathbf{a}_n \\ | & & | \end{bmatrix} \tag{3.29}$$

(recordemos que $\hat{\mathbf{e}}^k = \hat{\mathbf{e}}_k$). Es decir, si A es cuadrada y rango completo, de modo que tanto sus filas como sus columnas forman bases de $\mathbb{R}^n$, mirándola como colección de vectores fila, su inversa será una colección de vectores columna donde el vector columna $\mathbf{a}_i$ de la inversa es el elemento dual del vector fila $\mathbf{a}^i$. La prueba ya fue hecha en general debajo de (3.27).

De manera análoga, una dada matriz $n \times n$ de rango completo A, pensada como colección de columnas $\mathbf{a}_i$, tiene como inversa una matriz cuya fila i-ésima es el elemento $\left(\mathbf{a}^i\right)^{\top}$.

De (3.28) y (3.29) vemos que (3.17) y (3.18), para $j = n$, implican que:

$$\sum_{i=1}^{n} \mathbf{a}^i \, \mathbf{a}_i^{\top} = I \tag{3.30}$$

Recordemos que la base dual de la base dual de una base dada es la base original, por lo que se pueden intercambiar sub y super indices en (3.30).

Las siguientes propiedades de la matriz inversa son triviales a partir de lo visto:

$$\begin{aligned} \left(A^{-1}\right)^{-1} &= A \\ (AB)^{-1} &= B^{-1}A^{-1} \\ \left(A^{-1}\right)^{\top} &= \left(A^{\top}\right)^{-1} \end{aligned} \tag{3.31}$$

3.4 Pseudo inversa, o proyectores

Volvamos a una matriz cuadrada A que no es de rango completo, como en (3.24), actuando sobre las bases (3.22-3.23) y resultando en (3.25-3.26). Recordemos que $\left\{\mathbf{v}_{2,j+1}, \ldots, \mathbf{v}_{2,n}\right\}$ es la base de $\mathcal{N}(A)$ y $\left\{\mathbf{v}_{1,1}, \ldots, \mathbf{v}_{1,j}\right\}$ es la base de V_1, que $V_1 = V_2^{\perp}$, que $W_2 = W_1^{\perp}$, y que $\dim V_1 = \dim W_1$. Notar que como A es cuadrada, $V = W$, pero el subespacio V_1 no es necesariamente igual al subespacio W_1, ellos simplemente tienen la misma dimensión.

Llamamos "pseudo inversa" de (3.24) a la matriz

$$\tilde{A}^{-1} = \sum_{i=1}^{j} \mathbf{v}_{1,i} \left(\mathbf{w}^{1,i}\right)^{\top} \tag{3.32}$$

Por qué ese nombre? Consideremos el efecto de la matriz $\tilde{A}^{-1}A$ sobre un vector cualquiera. Dado que la base (3.22) es completa y linealmente independiente, todo vector $\mathbf{x} \in V$ se puede expresar de manera única como combinación lineal de los vectores de ella,

$$\mathbf{x} = \sum_{r=1}^{j} x^r \mathbf{v}_{1,r} + \sum_{r=j+1}^{n} x^r \mathbf{v}_{2,r} \tag{3.33}$$

por lo que, de (3.32) y (3.24):

$$\begin{aligned}
\tilde{A}^{-1}A\mathbf{x} &= \left(\sum_{i=1}^{j} \mathbf{v}_{1,i} \left(\mathbf{w}^{1,i}\right)^{\top}\right)\left(\sum_{k=1}^{j} \mathbf{w}_{1,k} \left(\mathbf{v}^{1,k}\right)^{\top}\right)\left(\sum_{r=1}^{j} x^r \mathbf{v}_{1,r} + \sum_{r=j+1}^{n} x^r \mathbf{v}_{2,r}\right) && (3.34)\\
&= \left(\sum_{i=1}^{j} \mathbf{v}_{1,i} \left(\mathbf{w}^{1,i}\right)^{\top}\right)\left(\sum_{k=1}^{j} \mathbf{w}_{1,k} x^k\right) && (3.35)\\
&= \sum_{i=1}^{j} x^i \mathbf{v}_{1,i} && (3.36)
\end{aligned}$$

para pasar de (3.34) a (3.35) usamos $\left(\mathbf{v}^{1,k}\right)^{\top} \mathbf{v}_{a,r} = \delta_a^1 \delta_r^k$. Y para pasar de (3.35) a (3.36) usamos $\left(\mathbf{w}^{1,i}\right)^{\top} \mathbf{w}_{1,k} = \delta_k^i$. Como se ve de la expansión (3.33), (3.36) es la proyección de $\mathbf{x}$ en V_1. De modo que si $\mathbf{x} \in V_1$, $\tilde{A}^{-1}A$ actúa como la matriz identidad. De ahí el nombre de $\tilde{A}^{-1}$ como pseudo inversa.

De manera análoga, dado que la base (3.23) es completa y linealmente independiente, todo vector $\mathbf{x} \in V$ se puede expresar de manera única como combinación lineal de los vectores de ella,

$$\mathbf{x} = \sum_{r=1}^{j} y^r \mathbf{w}_{1,r} + \sum_{r=j+1}^{n} y^r \mathbf{w}_{2,r} \tag{3.37}$$

por lo que, de (3.32) y (3.24):

$$A\tilde{A}^{-1}\mathbf{x} = \left(\sum_{i=1}^{j} \mathbf{w}_{1,i} \left(\mathbf{v}^{1,i}\right)^{\top}\right)\left(\sum_{k=1}^{j} \mathbf{v}_{1,k} \left(\mathbf{w}^{1,k}\right)^{\top}\right)\left(\sum_{r=1}^{j} y^r \mathbf{w}_{1,r} + \sum_{r=j+1}^{n} y^r \mathbf{w}_{2,r}\right) \tag{3.38}$$

$$= \left(\sum_{i=1}^{j} \mathbf{w}_{1,i} \left(\mathbf{v}^{1,i}\right)^{\top}\right)\left(\sum_{k=1}^{j} \mathbf{v}_{1,k} y^k\right) \tag{3.39}$$

$$= \sum_{i=1}^{j} y^i \mathbf{w}_{1,i} \tag{3.40}$$

Si $\mathbf{x} \in W_1$, $A\tilde{A}^{-1}$ actúa como la matriz identidad. De ahí el nombre de $\tilde{A}^{-1}$ como pseudo inversa.

Es importante recordar que si bien $\dim V_1 = \dim W_1$, W_1 es en general un subespacio de V diferente que V_1, por lo que en general $A\tilde{A}^{-1} \neq \tilde{A}^{-1}A$.

Para desarrollar el concepto de pseudoinversa supusimos que la matiz $A \in \mathbb{R}^{n\times n}$ es cuadrada, por lo que $V = W$, aunque no de rango completo. Pero mirando el desarrollo, desde la ecuación (3.32) hasta (3.40), notamos que todo funciona igual si $A \in \mathbb{R}^{m\times n}$, es decir, con $V = \mathbb{R}^n$ y $W = \mathbb{R}^m$ y con bases

$$\left\{\mathbf{v}_{1,1}, \ldots, \mathbf{v}_{1,j}, \mathbf{v}_{2,j+1}, \ldots, \mathbf{v}_{2,n}\right\} \tag{3.41}$$

$$\left\{\mathbf{w}_{1,1}, \ldots, \mathbf{w}_{1,j}, \mathbf{w}_{2,j+1}, \ldots, \mathbf{w}_{2,m}\right\} \tag{3.42}$$

Como vimos en la sección 2, sigue siendo el caso que $\mathcal{N}(A) = V_2$, $V_1 = V_2^{\perp}$, $\dim V_1 = \dim W_1$ (aunque ahora esos subespacios están sumergidos en espacios diferentes), y $W_2 = W_1^{\perp}$.

Resumiendo, $A\colon \mathbb{R}^n \to \mathbb{R}^m$, $\tilde{A}^{-1}\colon \mathbb{R}^m \to \mathbb{R}^n$. $\tilde{A}^{-1}A\colon \mathbb{R}^n \to \mathbb{R}^n$, es igual a la identidad en V_1, y a la matriz cero en V_2, es decir que "proyecta" vectores de V sobre V_1. $A\tilde{A}^{-1}\colon \mathbb{R}^m \to \mathbb{R}^m$ es igual a la identidad en W_1, y a la matriz cero en W_2, es decir que "proyecta" vectores de W sobre W_1.

$\tilde{A}^{-1}A$ tiene $\dim V_1$ autovalores repetidos igual a 1, y $\dim V_2$ autovalores repetidos igual a 0. Todos los vectores de V_1 son autovectores correspondientes al autovalor 1 (se puede elegir una base de $\dim V_1$ autovectores ortonormales que spanean V_1 y lo mismo con V_2).

De manera análoga, $A\tilde{A}^{-1}$ tiene $\dim W_1 = \dim V_1$ autovalores repetidos igual a 1, y $\dim W_2$ autovalores repetidos igual a 0. Todos los vectores de W_1 son autovectores correspondientes al autovalor 1 (se puede elegir una base de $\dim W_1$ autovectores ortonormales que spanean W_1 y lo mismo con W_2).

3.5 Proyecciones sobre una base dada de un subespacio

Consideremos una colección

$$\{\mathbf{x}_{1,1}, \ldots, \mathbf{x}_{1,j}\} \tag{3.43}$$

de j vectores de $\mathbb{R}^n$, $j < n$. Asumimos que son linealmente independientes pero no ortogonales.

Recordemos que el "span" de estos vectores es el conjunto de todas las posibles combinaciones lineales de ellos:

$$\text{span}\left(\left\{\mathbf{x}_{1,1}, \ldots \mathbf{x}_{1,j}\right\}\right) = \left\{\mathbf{v} : \mathbf{v} = \sum_{i=1}^{j} \alpha_i \mathbf{x}_{1,i}, \quad \alpha_i \in \mathbb{R}\right\} \tag{3.44}$$

Este conjunto es un subespacio de $\mathbb{R}^n$, al que llamamos V_1, de $\dim V_1 = j$ (por la asumida independencia lineal de los vectores $\mathbf{x}_i$). Llamemos V_2 al complemento ortogonal de V_1 en $\mathbb{R}^n$: $V_2 = V_1^{\perp}$, $\dim V_2 = n - j$.

Dado otro vector $\mathbf{y} \in \mathbb{R}^n$, buscamos al vector de V_1 más cercano a $\mathbf{y}$ en el sentido de ℓ_2. El teorema de Pitágoras nos indica que tal vector es la proyección de $\mathbf{y}$ sobre V_1:

$$\text{argmin}_{\mathbf{v} \in \text{span}\left(\left\{\mathbf{x}_{1,1}, \ldots, \mathbf{x}_{1,j}\right\}\right)} \|\mathbf{y} - \mathbf{v}\|_2 = \text{Proj}\,(\mathbf{y}; V_1) \tag{3.45}$$

Construimos la matriz D, cuyas columnas son los vectores $\{\mathbf{x}_{1,1}, \ldots, \mathbf{x}_{1,j}\}$, de modo que $D \in \mathbb{R}^{n \times j}$.

$$\underset{n \times j}{D} = \begin{bmatrix} | & & | \\ \mathbf{x}_{1,1} & \cdots & \mathbf{x}_{1,j} \\ | & & | \end{bmatrix} \tag{3.46}$$

Como mencionamos, dichos vectores son linealmente independientes, por lo que el rango, o imagen de D es V_1, de dimension j.

La siguiente ecuación resuelve el problema (3.45):

$$\text{Proj}(\mathbf{y}; V_1) = \text{argmin}_{\mathbf{v} \in W_1} \|\mathbf{v} - \mathbf{y}\|_2 = A\mathbf{y}, \tag{3.47}$$

$$\text{donde } \underset{n \times n}{A} = \underset{n \times j}{D} \left(\underset{j \times n}{D^{\top}} \underset{n \times j}{D}\right)^{-1} \underset{j \times n}{D^{\top}} \tag{3.48}$$

recordemos que $\mathbf{y} \in \mathbf{R}^{n \times 1}$, por lo que $A\mathbf{y} \in \mathbf{R}^{n \times 1}$. Y

$$\underset{j \times 1}{\alpha} = \left(\underset{j \times n}{D^{\top}} \underset{n \times j}{D}\right)^{-1} \underset{j \times n}{D^{\top}} \underset{n \times 1}{\mathbf{y}} \tag{3.49}$$

es un vector cuyas componentes son las coordenadas de $\text{Proj}(\mathbf{y}; V_1)$ en la base $\{\mathbf{x}_{1,1}, \ldots, \mathbf{x}_{1,j}\}$ (columnas de columnas de D).

Cuando D contiene solo una columna, $D = \mathbf{d} \in \mathbb{R}^{n \times 1}$, A en (3.48) se reduce trivialmente a la proyección de $\mathbf{y}$ en el subespacio spaneado por $\mathbf{d}$:

$$A = \mathbf{d}\left(\mathbf{d}^{\top}\mathbf{d}\right)^{-1}\mathbf{d}^{\top} = \frac{\mathbf{d}\mathbf{d}^{\top}}{\|\mathbf{d}\|^2} = \frac{\mathbf{d}}{\|\mathbf{d}\|}\frac{\mathbf{d}^{\top}}{\|\mathbf{d}\|} = \hat{\mathbf{d}}\hat{\mathbf{d}}^{\top} \tag{3.50}$$

donde $\hat{\mathbf{d}}$ es el vector de módulo unidad en la dirección de $\mathbf{d}$. (3.50) es un *proyector* en la dirección de $\mathbf{d}$, entonces, usando (1.8)

$$\text{Proj}(\mathbf{y}; \mathbf{d}) = \hat{\mathbf{d}}\left(\hat{\mathbf{d}}^{\top}\mathbf{y}\right) = \|\mathbf{y}\| \cos\theta_{\mathbf{d},\mathbf{y}} \hat{\mathbf{d}} \tag{3.51}$$

donde $\theta_{\mathbf{d},\mathbf{y}}$ es el ángulo entre $\mathbf{d}$ e $\mathbf{y}$, por lo que (3.51) es transparentemente la proyección de $\mathbf{y}$ en la dirección de $\mathbf{b}$, como habíamos anunciado.

Notar, por otra parte, que

$$\left(\mathbf{d}^\top \mathbf{d}\right)^{-1} \mathbf{d}^\top \mathbf{y} = \frac{\mathbf{d}^\top \mathbf{y}}{||\mathbf{d}||^2} \tag{3.52}$$

es la coordenada del vector $\text{Proj}(\mathbf{y}; \mathbf{d})$ de (3.51) en la base constituida por $\mathbf{d}$ (a diferencia de la base ortonormal constituida por $\hat{\mathbf{d}}$).

Este resultado es muy simple, y está descrito en cualquier libro de textos básico de álgebra lineal. Por otra parte, que la fórmula general (3.47-3.48) también genera la proyección de $\mathbf{y}$ en V_1, no es tan fácil de intuir en general con los métodos estándar. El resultado no es tan transparente como (3.51).

3.5.1 La proyección no es directa cuando la base no es ortonormal

La razón fundamental de la complicación es que, si bien los vectores $\{\mathbf{x}_{1,1}, \ldots, \mathbf{x}_{1,j}\}$ son linealmente independientes, no son ortogonales. Cuando una base no es ortogonal, la suma de las proyecciones individuales no es igual a la proyección en el subespacio que spanean.

Por ejemplo, consideremos el vector de $\mathbb{R}^3$ $\mathbf{y} = (1, 1, 1)^\top$. Su proyección sobre el plano (x, y) es obviamente $\text{Proj}(\mathbf{y};\ \text{plano}(x, y)) = (1, 1, 0)^\top$. Pero consideremos en este plano la base no ortogonal dada por $\mathbf{x}_1 = (1, 0, 0)^\top$ y $\mathbf{x}_2 = (1, 1, 0)^\top / \sqrt{2}$. Ambos tienen norma 1, pero no son ortogonales entre sí. Veamos las proyecciones de $\mathbf{y}$ sobre ellos y su suma:

$$\begin{aligned}
\text{Proj}(\mathbf{y}; \mathbf{x}_1) &= \mathbf{x}_1 \left(\mathbf{x}_1^\top \mathbf{y}\right) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \left[(1, 0, 0) \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right] = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} & (3.53) \\
\text{Proj}(\mathbf{y}; \mathbf{x}_2) &= \mathbf{x}_2 \left(\mathbf{x}_2^\top \mathbf{y}\right) = \frac{1}{2} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \left[(1, 1, 0) \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right] = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} & (3.54) \\
\text{Proj}(\mathbf{y}; \mathbf{x}_1) + \text{Proj}(\mathbf{y}; \mathbf{x}_2) &= \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} \neq \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \text{Proj}\left(\mathbf{y}; \text{span}\left(\{\mathbf{x}_1, \mathbf{x}_2\}\right)\right) & (3.55)
\end{aligned}$$

Si $\mathbf{x}_1$ y $\mathbf{x}_2$ hubieran sido ortogonales, esta suma sí hubiera funcionado. Por ejemplo si $\mathbf{x}'_1$ =

$(1,-1,0)^\top/\sqrt{2}$ y $\mathbf{x}_2' = (1,1,0)^\top/\sqrt{2}\}$ (comprobar que son ortonormales):

$$\mathrm{Proj}(\mathbf{y};\mathbf{x}_1') = \mathbf{x}_1'\left((\mathbf{x}_1')^\top\mathbf{y}\right) = \frac{1}{2}\begin{pmatrix}1\\-1\\0\end{pmatrix}\left[(1,-1,0)\begin{pmatrix}1\\1\\1\end{pmatrix}\right] = \begin{pmatrix}0\\0\\0\end{pmatrix} \tag{3.56}$$

$$\mathrm{Proj}(\mathbf{y};\mathbf{x}_2') = \mathbf{x}_2'\left((\mathbf{x}_2')^\top\mathbf{y}\right) = \frac{1}{2}\begin{pmatrix}1\\1\\0\end{pmatrix}\left[(1,1,0)\begin{pmatrix}1\\1\\1\end{pmatrix}\right] = \begin{pmatrix}1\\1\\0\end{pmatrix} \tag{3.57}$$

$$\mathrm{Proj}(\mathbf{y};\mathbf{x}_1') + \mathrm{Proj}(\mathbf{y};\mathbf{x}_2') = \begin{pmatrix}1\\1\\0\end{pmatrix} = \mathrm{Proj}\left(\mathbf{y};\mathrm{span}\left(\{\mathbf{x}_1',\mathbf{x}_2'\}\right)\right) \tag{3.58}$$

En abstracto, si expresamos al vector $\mathbf{y}$ en una base ortonormal, esto es, una base $\{\mathbf{e}_1,\ldots,\mathbf{e}_n\}$:

$$\mathbf{y} = \sum_{i=1}^{n} y_i\mathbf{e}_i, \quad \text{donde} \quad \mathbf{e}_i^\top\mathbf{e}_j = \delta_{ij} \tag{3.59}$$

obtenemos las coordenadas y_i de $\mathbf{y}$ en esta base como:

$$\mathbf{e}_j^\top\mathbf{y} = \sum_{i=1}^{n} y_i\mathbf{e}_j^\top\mathbf{e}_i = y_j \tag{3.60}$$

por lo que la matriz $\sum_{k=1}^{j}\mathbf{e}_k\mathbf{e}_k^\top$ proyecta sobre el subespacio spaneado por $\{\mathbf{e}_1,\ldots,\mathbf{e}_j\}$:

$$\left(\sum_{k=1}^{j}\mathbf{e}_k\mathbf{e}_k^\top\right)\mathbf{y} = \sum_{k=1}^{j}\sum_{i=1}^{n} y_i\left(\mathbf{e}_k\mathbf{e}_k^\top\right)\mathbf{e}_i = \sum_{k=1}^{j}\sum_{i=1}^{n} y_i\mathbf{e}_k\left(\mathbf{e}_k^\top\mathbf{e}_i\right) = \sum_{k=1}^{j} y_k\mathbf{e}_k \tag{3.61}$$

Si, por el contrario, la base de vectores linealmente independientes $\{\mathbf{x}_1,\ldots,\mathbf{x}_n\}$ no es ortogonal, si bien sigue siendo cierto que podemos expresar a cualquier vector $\mathbf{y}$ de manera única en dicha base:

$$\mathbf{y} = \sum_{i=1}^{n} y_i'\mathbf{x}_i \tag{3.62}$$

como ya no es cierto que $\mathbf{x}_j^\top\mathbf{x}_j = \delta_{ji}$, no es automática la extracción de las coordenadas y_i' en esta base:

$$\mathbf{x}_j^\top\mathbf{y} = \sum_{i=1}^{n} y_i\left(\mathbf{x}_j^\top\mathbf{x}_i\right) \tag{3.63}$$

como el paréntesis del lado derecho no es δ_{ji}, no se puede proyectar de manera análoga a (3.61).

Una alternativa para proyectar a $\mathbf{y}$ en es subespacio spaneado por $\{\mathbf{x}_{1,1},\ldots,\mathbf{x}_{1,j}\}$ sería usar el proceso de Gram-Schmidt para generar una base ortonormal que spanee el mismo subespacio. Esto funciona, pero luego, aplicando los procedimientos (3.60-3.61), obtendríamos las coordenadas en esta base de Gram-Schmidt.

Resulta que, como vamos a ver más adelante, para muchas aplicaciones, por ejemplo regresiones, no sólo necesitamos obtener la proyección de $\mathbf{y}$ en es subespacio spaneado por $\{\mathbf{x}_1,\ldots,\mathbf{x}_j\}$, sino que además necesitamos sus coordenadas en esa misma base. Requeriría otro proceso adicional extraer las coordenadas en esta base a partir de las coordenadas en la base de Gram-Schmidt.

3.5.2 La proyección es directa usando bases duales aunque la base no sea ortonormal

Es una prueba del poder del formalismo aprendido, verificar la trivialidad con la que se logra este objetivo usando bases duales que, como dijimos en (2.55), devuelven la simpleza y transparencia de las bases ortonormales a bases que no lo son. De hecho, esencialmente ya resolvimos este problema en (3.7-3.10).

Primero veamos cómo construimos dichas bases: en el espíritu de lo hecho en la sección 3.1, a partir de los vectores $\{\mathbf{x}_{1,1}, \ldots, \mathbf{x}_{1,j}\}$ dados en (3.43) construimos la base de $\mathbb{R}^n$

$$\left\{\mathbf{x}_{1,1}, \ldots, \mathbf{x}_{1,j}, \mathbf{x}_{2,j+1}, \ldots, \mathbf{x}_{2,n}\right\} \tag{3.64}$$

donde $\{\mathbf{x}_{1,1}, \ldots, \mathbf{x}_{1,j}\}$ es base de V_1 y $\{\mathbf{x}_{2,j+1}, \ldots, \mathbf{x}_{2,n}\}$ es base del complemento ortogonal $V_2 = V_1^\perp$ en $\mathbb{R}^n$. Esto se puede hacer en la práctica con la función **scipy.linalg.null_space()** aplicada a $D^\top$, con D en (3.46).

En la base (3.64), porque $V_2 = V_1^\perp$, $\mathbf{x}_{1,i}^\top \mathbf{x}_{2,k} = 0$ para todo $i = 1, \ldots, j$ y todo $k = j+1, \ldots, n$, pero recordemos que los $\{\mathbf{x}_{1,1}, \ldots, \mathbf{x}_{1,j}\}$, dados en (3.43), no son ortogonales entre sí. A partir de (3.64) construimos la base dual

$$\left\{\mathbf{x}^{1,1}, \ldots, \mathbf{x}^{1,j}, \mathbf{x}^{2,j+1}, \ldots, \mathbf{x}^{2,n}\right\} \tag{3.65}$$

donde $\{\mathbf{x}^{1,1}, \ldots, \mathbf{x}^{1,j}\}$ es base de V_1 y $\{\mathbf{x}^{2,j+1}, \ldots, \mathbf{x}^{2,n}\}$ es base de V_2. En la práctica, la base (3.65) se puede calcular con el procedimiento explicado en la sección 2.3.

Con las bases (3.64) y (3.65), estamos en una situación idéntica a la estudiada en la sección 3.1, por lo que copiamos la solución. Como la base (3.64), que contiene a los vectores (3.43), es completa, existe una expansion del vector $\mathbf{y}$ de (3.45) como en (3.5):

$$\mathbf{y} = \sum_{i=1}^{j} y^{1,i} \mathbf{x}_{1,i} + \sum_{k=j+1}^{n} y^{2,k} \mathbf{x}_{2,k} \tag{3.66}$$

La proyección deseada de $\mathbf{y}$ en V_1 (ver (3.45)) es la primera suma en esta expresión. El problema es cómo encontrar las coordenadas $y^{1,i}$. La solución, que reescribimos, está dada en (3.7):

$$y^{1,r} = \left(\mathbf{x}^{1,r}\right)^\top \mathbf{y} \tag{3.67}$$

La similitud entre (3.67) y (3.60) muestra que, efectivamente, las bases duales devuelven la simpleza y transparencia de las bases ortonormales a bases que no lo son.

Con las bases duales también podemos trivialmente escribir un proyector análogo al que hicimos en (3.61) para bases ortogonales:

$$A = \sum_{r=1}^{j} \mathbf{x}_{1,r} \left(\mathbf{x}^{1,r}\right)^\top \tag{3.68}$$

Verifiquemos que (3.68) logra no sólo la proyección de $\mathbf{y}$ sino que además lo hace en la base deseada:

$$A\mathbf{y} = \sum_{r=1}^{j} \sum_{i=1}^{j} y^{1,i} \mathbf{x}_{1,r} \left(\mathbf{x}^{1,r}\right)^\top \mathbf{x}_{1,i} + \sum_{r=1}^{j} \sum_{k=j+1}^{n} y^{2,k} \mathbf{x}_{1,r} \left(\mathbf{x}^{1,r}\right)^\top \mathbf{x}_{2,k} = \sum_{r=1}^{j} y^{1,r} \mathbf{x}_{1,r} \tag{3.69}$$

El problema (3.45) está entonces resuelto.

Comparando el proyector (3.61) en la base ortogonal, $\sum_{k=1}^{j} \mathbf{e}_k \mathbf{e}_k^\top$, con el proyector (3.68) en la base no ortogonal (3.64) vemos una vez más el poder del formalismo dual. (3.68) se reduce a (3.61) si la base (3.64) fuera ortogonal, porque la base dual coincide con la original si esta es ortogonal.

Volvamos al ejemplo de la sección 3.5.1.

Recordemos que los vectores $\mathbf{x}_1$ y $\mathbf{x}_2$, que ahora pasamos a llamar $\mathbf{x}_{1,1}$ y $\mathbf{x}_{1,2}$, son $\mathbf{x}_{1,1} = (1, 0, 0)^\top$ y $\mathbf{x}_{1,2} = (1, 1, 0)^\top$ (no dividimos este vector por $\sqrt{2}$, como sí lo hicimos en 3.5.1, para mostrar que el formalismo funciona independientemente de la norma de los vectores), forman la base de V_1.

Podemos elegir la base de $V_2 = V_1^\perp$ como $\mathbf{x}_{2,3} = (0, 0, 1)^\top$. Por supuesto que hay arbitrariedad acá, porque lo único que se requiere de $\mathbf{x}_{2,3}$ es que sea ortogonal a V_1, y cualquier vector de la forma $(0, 0, c)^\top$, para $c \in \mathbb{R}$, cumple con dicha condición. Pero por simplicidad elegimos $c = 1$. Entonces la base $\{\mathbf{x}_{1,1}, \mathbf{x}_{1,2}, \mathbf{x}_{2,3}\}$ (3.64) es

$$\mathbf{x}_{1,1} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{x}_{1,2} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{x}_{2,3} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \tag{3.70}$$

la base dual (3.65) es

$$\mathbf{x}^{1,1} = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \quad \mathbf{x}^{1,2} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{x}^{2,3} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \tag{3.71}$$

el lector puede comprobar estas bases cumplen las condiciones (3.3-3.4).

Queremos proyectar un vector arbitrario de $\mathbb{R}^3$:

$$\mathbf{y} = \begin{pmatrix} a \\ b \\ c \end{pmatrix} \tag{3.72}$$

en V_1. En este caso V_1 es el plano (x, y), así que su proyección es obviamente $\mathrm{Proj}(\mathbf{y}; V_1) = (a, b, 0)^\top$. Pero no queremos sólo proyectar y en V_1, queremos hacerlo conociendo las coordenadas en la base $\{\mathbf{x}_{1,1}, \mathbf{x}_{1,2}\}$. (3.67) nos indica que

$$y^{1,1} = \left(\mathbf{x}^{1,1}\right)^\top \mathbf{y} = (1, -1, 0) \begin{pmatrix} a \\ b \\ c \end{pmatrix} = a - b \tag{3.73}$$

$$y^{1,2} = \left(\mathbf{x}^{1,2}\right)^\top \mathbf{y} = (0, 1, 0) \begin{pmatrix} a \\ b \\ c \end{pmatrix} = b \tag{3.74}$$

de modo que

$$y^{1,1}\mathbf{x}_{1,1} + y^{1,2}\mathbf{x}_{1,2} = (a-b)\begin{pmatrix}1\\0\\0\end{pmatrix} + b\begin{pmatrix}1\\1\\0\end{pmatrix} = \begin{pmatrix}a\\b\\0\end{pmatrix} = \mathrm{Proj}(\mathbf{y}; V_1) \quad \checkmark \tag{3.75}$$

entonces (3.73) y (3.74) son las coordenadas correctas.

La matriz A en (3.68) es:

$$\begin{aligned}
A &= \sum_{r=1}^{2} \mathbf{x}_{1,r}\left(\mathbf{x}^{1,r}\right)^{\top} \\
&= \begin{pmatrix}1\\0\\0\end{pmatrix}(1,-1,0) + \begin{pmatrix}1\\1\\0\end{pmatrix}(0,1,0) \\
&= \begin{pmatrix}1 & -1 & 0\\0 & 0 & 0\\0 & 0 & 0\end{pmatrix} + \begin{pmatrix}0 & 1 & 0\\0 & 1 & 0\\0 & 0 & 0\end{pmatrix} \\
&= \begin{pmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 0\end{pmatrix}
\end{aligned} \tag{3.76}$$

de modo que

$$A\mathbf{y} = \begin{pmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 0\end{pmatrix}\begin{pmatrix}a\\b\\c\end{pmatrix} = \begin{pmatrix}a\\b\\0\end{pmatrix} = \mathrm{Proj}(\mathbf{y}; V_1) \tag{3.77}$$

acorde con (3.69).

3.5.3 Iluminando la solución $A = D\left(D^{\top} D\right)^{-1} D^{\top}$

La matriz A en (3.48) no se parece a la matriz A en (3.68). Más aún, el significado de (3.48) es completamente opaco, mientras que (3.68) es transparentemente un proyector en el subespacio V_1 spaneado por los vectores $\{\mathbf{x}_{1,1}, \ldots, \mathbf{x}_{1,j}\}$ de (3.64). En esta sección demostramos que (3.48) y (3.68) son la misma matriz.

Expresemos la matriz D en (3.46) como suma de productos outer:

$$\underset{n\times j}{D} = \begin{bmatrix} | & & | \\ \mathbf{x}_{1,1} & \cdots & \mathbf{x}_{1,j} \\ | & & | \end{bmatrix} = \sum_{r=1}^{j} \mathbf{x}_{1,r}\hat{\mathbf{e}}_r^{\top}, \qquad \underset{j\times n}{D^{\top}} = \begin{bmatrix} -\mathbf{x}_{1,1}^{\top}- \\ \vdots \\ -\mathbf{x}_{1,j}^{\top}- \end{bmatrix} = \sum_{s=1}^{j} \hat{\mathbf{e}}_s \mathbf{x}_{1,s}^{\top} \tag{3.78}$$

donde $\{\hat{\mathbf{e}}_i\}$, $i = 1, \ldots, j$ es la base canónica de $\mathbb{R}^d$, que escribimos siempre con subíndices porque, ya que es ortonormal, su base dual coincide consigo misma. Estamos usando las formas (2.39) y (2.40) de escribir una matriz. Notar que la segunda expresión sale de la primera recordando que la transpuesta de una suma es la suma de las transpuestas y que la transpuesta de un producto es el producto de las transpuestas intercambiando posiciones.

De (3.78),

$$D^\top D = \left(\sum_{s=1}^{j} \hat{\mathbf{e}}_s \mathbf{x}_{1,s}^\top\right)\left(\sum_{r=1}^{j} \mathbf{x}_{1,r} \hat{\mathbf{e}}_r^\top\right) = \sum_{s,r=1}^{j} \hat{\mathbf{e}}_s \mathbf{x}_{1,s}^\top \mathbf{x}_{1,r} \hat{\mathbf{e}}_r^\top \tag{3.79}$$

De (3.79) no es difícil adivinar que

$$\left(D^\top D\right)^{-1} = \sum_{u,v=1}^{j} \hat{\mathbf{e}}_u \left(\mathbf{x}^{1,u}\right)^\top \mathbf{x}^{1,v} \hat{\mathbf{e}}_v^\top \tag{3.80}$$

lo verificamos así:

$$\begin{aligned}
\left(D^\top D\right)^{-1}\left(D^\top D\right) &= \sum_{u,v,s,r=1}^{j} \hat{\mathbf{e}}_u \left(\mathbf{x}^{1,u}\right)^\top \mathbf{x}^{1,v} \underbrace{\hat{\mathbf{e}}_v^\top \hat{\mathbf{e}}_s}_{=\delta_{vs}} \mathbf{x}_{1,s}^\top \mathbf{x}_{1,r} \hat{\mathbf{e}}_r^\top && (3.81)\\
&= \sum_{u,r=1}^{j} \hat{\mathbf{e}}_u \left(\mathbf{x}^{1,u}\right)^\top \underbrace{\left(\sum_{v=1}^{j} \mathbf{x}^{1,v} \mathbf{x}_{1,v}^\top\right)}_{=I_{V_1}} \mathbf{x}_{1,r} \hat{\mathbf{e}}_r^\top && (3.82)\\
&= \sum_{u,r=1}^{j} \hat{\mathbf{e}}_u \underbrace{\left(\mathbf{x}^{1,u}\right)^\top \mathbf{x}_{1,r}}_{=\delta_{ur}} \hat{\mathbf{e}}_r^\top && (3.83)\\
&= \sum_{u=1}^{j} \hat{\mathbf{e}}_u \hat{\mathbf{e}}_u^\top = \underset{j\times j}{I} && (3.84)
\end{aligned}$$

El único paso no trivial a primera vista es (3.82). Como vimos en (3.68), el paréntesis es un proyector sobre V_1 (o la matriz identidad restringida a ese subespacio, de ahí el nombre I_{V_1}), y como $\mathbf{x}_{1,r} \in V_1$, $\forall r = 1, \ldots, j$, $I_{V_1}\mathbf{x}_{1,r} = \mathbf{x}_{1,r}$. Se deja para el lector la comprobación de $\left(D^\top D\right)\left(D^\top D\right)^{-1} = I$.

Con (3.78) y (3.80),

$$\begin{aligned}
D\left(D^\top D\right)^{-1} D^\top &= \sum_{r,u,v,s=1}^{j} \mathbf{x}_{1,r} \underbrace{\hat{\mathbf{e}}_r^\top \hat{\mathbf{e}}_u}_{=\delta_{ru}} \left(\mathbf{x}^{1,u}\right)^\top \mathbf{x}^{1,v} \underbrace{\hat{\mathbf{e}}_v^\top \hat{\mathbf{e}}_s}_{=\delta_{vs}} \mathbf{x}_{1,s}^\top && (3.85)\\
&= \underbrace{\left(\sum_{r=1}^{j} \mathbf{x}_{1,r} \left(\mathbf{x}^{1,r}\right)^\top\right)}_{=I_{V_1}} \underbrace{\left(\sum_{s=1}^{j} \mathbf{x}^{1,s} \left(\mathbf{x}_{1,s}\right)^\top\right)}_{=I_{V_1}} && (3.86)\\
&= \sum_{r=1}^{j} \mathbf{x}_{1,r} \left(\mathbf{x}^{1,r}\right)^\top && (3.87)
\end{aligned}$$

en la última igualdad usamos la identidad $I_{V_1} I_{V_1} = I_{V_1}$, válida para todo proyector. (3.87), que es (3.48), es idéntica a (3.68), como queríamos probar. ✓

Recordemos que en (3.67) están dadas las coordenadas del vector proyección en la base de columnas de D (3.78). En (3.49) se afirma que esas coordenadas, ordenadas como las componentes de un vector $\in R^d$, vienen dadas por $\boldsymbol{\alpha} = \left(D^\top D\right)^{-1} D^\top \mathbf{y}$. Es fácil ver que esto tiene que ser así, porque si, como acabamos de probar,

$$\operatorname{Proj}(\mathbf{y}; V_1) = D \left(D^\top D\right)^{-1} D^\top \mathbf{y} \tag{3.88}$$

la forma (1.14) de ver el producto matriz-vector nos indica que $D \in \mathbb{R}^{n\times d}$, multiplicada por un vector $\boldsymbol{\alpha} \in \mathbb{R}^{d\times 1}$, resulta en una combinación lineal de las columnas de D, con coeficientes iguales a las componentes de $\boldsymbol{\alpha}$. Pero eso es precisamente el vector proyección $\operatorname{Proj}(\mathbf{y}; V_1)$ en (3.88). Por lo tanto

$$\boldsymbol{\alpha} = \left(D^\top D\right)^{-1} D^\top \mathbf{y} \tag{3.89}$$

es un vector cuyas componentes son las coordenadas de $\operatorname{Proj}(\mathbf{y}; V_1)$ en la base de columnas de D.

Ejercicio: Verificar explícitamente (al estilo de la prueba (3.85-3.87)), que las componentes del vector $\boldsymbol{\alpha}$ en (3.89) vienen dadas por (3.67) (esto es, por $y^{1,r} = \left(\mathbf{x}^{1,r}\right)^\top \mathbf{y}$).

Veamos el ejemplo analizado en 3.5.2 con $\mathbf{x}_{1,1} = (1,0,0)^\top$ y $\mathbf{x}_{1,2} = (1,1,0)^\top$:

$$D = \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \tag{3.90}$$

entonces

$$D^\top D = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}, \qquad \left(D^\top D\right)^{-1} = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} \tag{3.91}$$

por lo tanto

$$\begin{aligned} D\left(D^\top D\right)^{-1} D^\top &= \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \end{aligned} \tag{3.92}$$

igual a (3.76), como habíamos probado en (3.85-3.85), pero definitivamente con un procedimiento más opaco.

Si, como ocurre en aplicaciones a regresiones (ver sección 4), lo que realmente queremos son las coordenadas en la base de las columnas de D en (3.90), la solución viene dada por (3.73) y (3.74), o, para el caso general, por (3.67).

4 Regresiones

4.1 La regresión es una proyección en un subespacio determinado por los datos

En problemas de regresión, típicamente tenemos una variable, a la que llamamos y, a la que queremos "explicar" en términos de otras $j-1$ variables $x_1, \ldots, x_{j-1}$, con un modelo lineal:

$$y = \alpha_0 + x_1\alpha_1 + \ldots + x_{j-1}\alpha_{j-1} = x_0\alpha_0 + x_1\alpha_1 + \ldots + x_{j-1}\alpha_{j-1} \tag{4.1}$$

donde α_0 es la ordenada al origen. En la segunda igualdad introducimos una variable x_0 ficticia que siempre va a valer $x_0 = 1$.

Usualmente contamos con datos, esto es, valores empíricos para las variables $x_1, \ldots, x_{j-1}$, y para el correspondiente valor de y_{emp}. Supongamos que contamos con n conjuntos de datos, a cada uno de esos conjuntos los llamamos muestras. Ordenamos los valores de las xs en una matriz $D \in \mathbb{R}^{n\times j}$ y los de las y_{emp} en un vector $\mathbf{y}_{\text{emp}} \in \mathbb{R}^{n\times 1}$:

$$\underset{n\times j}{D} = \begin{pmatrix} x_{0;1,0} & x_{0;1,1} & \cdots & x_{0;1,j-1} \\ x_{1;1,0} & x_{1;1,1} & \cdots & x_{1;1,j-1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n-1;1,0} & x_{n-1;1,1} & \cdots & x_{n-1;1,j-1} \end{pmatrix}, \qquad \underset{n\times 1}{\mathbf{y}_{\text{emp}}} = \begin{pmatrix} y_{\text{emp},0} \\ y_{\text{emp},1} \\ \vdots \\ y_{\text{emp},n-1} \end{pmatrix} \tag{4.2}$$

En la matriz D, el primer índice numera a la muestra, va de 0 a $n-1$, porque asumimos que tenemos n muestras. El segundo índice, que como se ve, por ahora son todos 1, se va a explicar más adelante, y el tercer índice indica la variable explicativa $k = 0, \ldots, j-1$. Como dijimos, la variable ficticia $k = 0$ siempre toma el valor 1, es decir que toda la primera columna de D tiene 1s. Notar que los índices ahora corren de 0 a $j-1$, y de 0 a $n-1$, a diferencia del resto del paper en el que corrían de 1 a j, o de 1 a n.

Para la muestra i, la "predicción" del valor de la variable y en (4.1) es

$$y_{\text{pred},i} = x_{i;1,0}\alpha_0 + x_{i;1,1}\alpha_1 + \ldots + x_{i;1,j-1}\alpha_{j-1}, \qquad i = 0, \ldots, n-1 \tag{4.3}$$

Podemos "vectorizar" las predicciones de las distintas muestras:

$$\mathbf{y}_{\text{pred}} = \begin{pmatrix} y_{\text{pred},0} \\ y_{\text{pred},1} \\ \vdots \\ y_{\text{pred},n-1} \end{pmatrix} = \begin{pmatrix} x_{0;1,0}\alpha_0 + \ldots + x_{0;1,j-1}\alpha_{j-1} \\ x_{1;1,0}\alpha_0 + \ldots + x_{1;1,j-1}\alpha_{j-1} \\ \vdots \\ x_{n-1;1,0}\alpha_0 + \ldots + x_{n-1;1,j-1}\alpha_{j-1} \end{pmatrix} = D\boldsymbol{\alpha} \tag{4.4}$$

donde $\boldsymbol{\alpha} \in R^{j\times 1}$ es el vector de los parámetros α_i:

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_{j-1} \end{pmatrix} \tag{4.5}$$

Notar que el vector $\mathbf{y}_{\text{pred}}$ en (4.4), pensando en el producto matriz vector de la forma (1.14), se puede pensar como una combinación lineal de los vectores columna de D:

$$\mathbf{y}_{\text{pred}} = D\boldsymbol{\alpha} = \sum_{i=0}^{j-1} \mathbf{x}_{1,i}\alpha_i \tag{4.6}$$

donde

$$\mathbf{x}_{1,i} = \begin{pmatrix} x_{0;1,i} \\ x_{1;1,i} \\ \vdots \\ x_{n-1;1,i} \end{pmatrix}, \qquad i = 0, \ldots, j-1 \tag{4.7}$$

Los "errores" cometidos por el modelo (4.1) para un dado vector de parámetros $\boldsymbol{\alpha}$ también se pueden vectorizar en un vector **err**:

$$\mathbf{err} = \mathbf{y}_{\text{emp}} - \mathbf{y}_{\text{pred}} = \begin{pmatrix} y_{\text{emp},0} - y_{\text{pred},0} \\ y_{\text{emp},1} - y_{\text{pred},1} \\ \vdots \\ y_{\text{emp},n-1} - y_{\text{pred},n-1} \end{pmatrix} = \begin{pmatrix} y_{\text{emp},0} - \left(x_{0;1,0}\alpha_0 + \ldots + x_{0;1,j-1}\alpha_{j-1}\right) \\ y_{\text{emp},1} - \left(x_{1;1,0}\alpha_0 + \ldots + x_{1;1,j-1}\alpha_{j-1}\right) \\ \vdots \\ y_{\text{emp},j-1} - \left(x_{n-1;1,0}\alpha_0 + \ldots + x_{n-1;1,j-1}\alpha_{j-1}\right) \end{pmatrix} \tag{4.8}$$

El método de regresiones consiste en elegir los parámetros (4.5) de modo de minimizar la suma del cuadrado de los errores de predicción. Esto es, definiendo una funcion "error", o "costo":

$$\mathrm{C}(\alpha_0, \ldots, \alpha_{j-1}) = \sum_{i=0}^{j-1} \left[y_{\text{emp},i} - \left(x_{i;1,0}\alpha_0 + \ldots + x_{i;1,j-1}\alpha_{j-1}\right)\right]^2 \tag{4.9}$$

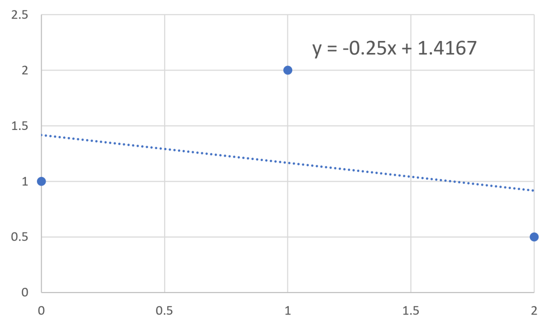
$$\text{objetivo:} \quad \min_{\{\alpha_0,\ldots,\alpha_{j-1}\}} \mathrm{C} \tag{4.10}$$

Comparando (4.9) con (4.8), es fácil ver que la función costo es el cuadrado de la norma ℓ_2 del vector **err**. Este vector, a su vez, es la resta vectorial $\mathbf{y}_{\text{emp}} - \mathbf{y}_{\text{pred}}$. Pero como vimos en (4.6), para todos los posibles valores de los parámetros α_i, $\mathbf{y}_{\text{pred}}$ cubre todo el subespacio V_1 spaneado por los vectores $\mathbf{x}_{1,i}$ de (4.7), columnas de la matriz D en (4.2).

De modo que el objetivo (4.10) es equivalente a encontrar el vector del subespacio V_1 más cercano al vector $\mathbf{y}_{\text{emp}} \in R^{n\times 1}$, lo que va a minimizar el cuadrado del módulo ℓ_2 del vector **err**. Esto corresponde a la proyección ortogonal de $\mathbf{y}_{\text{emp}}$ en V_1. Pero ese es exactamente el problema que resolvimos en la sección 3.5, donde los vectores $\mathbf{x}_{1,i}$ en (4.6) juegan el rol de los vectores en (3.43) en dicha sección!

Esta es la razón del segundo subíndice, siempre igual a 1, en los valores empíricos $x_{i;1,k}$ en (4.2), preparando el terreno para construir la base (3.64), que incluye a los vectores $\left\{\mathbf{x}_{2,j}, \ldots, \mathbf{x}_{1,n-1}\right\}$ del complemento ortogonal $V_2 = V_1^{\perp}$ de $\mathbb{R}^n$.

Por lo tanto, las matrices (3.48), o (3.68), generan la proyección ortogonal de $\mathbf{y}_{\text{emp}}$ en V_1: $A\mathbf{y}_{\text{emp}} = \mathbf{y}_{\text{pred}}$. Y la solución del problema de regresión, esto es, encontrar los parámetros α_i óptimos que



(a) Solución en Excel.

```
"""Regresiones""" # Ver ejemplo Excel en el paper
x10 = np.array([[1],[1],[1]])
x11 = np.array([[0],[1],[2]])
D = np.concatenate((x10, x11), axis=1)
x22 = scipy.linalg.null_space(np.transpose(D))
```

```
x1122 = np.concatenate((x11,x22), axis=1)
x10d = scipy.linalg.null_space(np.transpose(x1122))
x10d /= np.dot(np.transpose(x10),x10d)

x1022 = np.concatenate((x10,x22), axis=1)
x11d = scipy.linalg.null_space(np.transpose(x1022))
x11d /= np.dot(np.transpose(x11),x11d)

x22d = x22
```

```
y = np.array([[1],[2],[0.5]])
y_x10 = np.dot(np.transpose(x10d),y)
y_x11 = np.dot(np.transpose(x11d),y)
print("y_x10 =", y_x10)
print("y_x11 =", y_x11)
```

```
y_x10 = [[1.41666667]]
y_x11 = [[-0.25]]
```

(b) Solución con bases duales.

Figure 5: Regresión. Datos: $(0, 1), (1, 2), (2, 0.5)$

logran el objetivo (4.10), son las coordenadas de $\mathbf{y}_{\text{pred}}$ en la base de determinada por los vectores columna de D en (4.2). Estas coordenadas vienen dadas por (3.67), que son las componentes del vector $\boldsymbol{\alpha}$ en (3.49). ✓

4.1.1 Ejercicio

Supongamos que queremos "explicar" la variable y, teniendo los siguientes datos "empíricos": $(0, 1), (1, 2), (2, 0.5)$, y queremos encontrar la recta óptima $y = x_0\alpha_0 + x_1\alpha_1$, donde recordemos que $x_0 = 1$. En la figura 5a tenemos la solución en Excel del problema de regresión. En la figura 5b tenemos la solución con bases duales. La matriz D y el vector $\mathbf{y}_{\text{emp}}$ de (4.2) son:

$$D = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix}, \qquad \mathbf{y}_{\text{emp}} = \begin{pmatrix} 1 \\ 2 \\ 0.5 \end{pmatrix} \tag{4.11}$$

1. Graficar la solución en 3-D, visualizando la proyección de $\mathbf{y}_{\text{emp}}$ sobre el subespacio spaneado por las columnas de D.
2. Si llamamos $\mathbf{x}_{1,0} = (1, 1, 1)^\top$ y $\mathbf{x}_{1,1} = (0, 1, 2)^\top$ (primera y segunda columna de D en (4.11)), y $\mathbf{x}_{2,2}$ a un vector de norma 1 ortogonal a estos dos (ver figura 5b).
   (a) Construir en Python la matrix $\mathbf{x}_{1,0}\left(\mathbf{x}^{1,0}\right)^\top$ y verificar que es un proyector sobre $\mathbf{x}_{1,0}$. Repetir el ejercicio con el proyector sobre $\mathbf{x}_{1,1}$ y el proyector sobre $\mathbf{x}_{2,2}$.
   (b) V_1 es el subespacio spaneado por $\{\mathbf{x}_{1,0}, \mathbf{x}_{1,1}\}$, y V_2 es el subespacio spaneado por $\{\mathbf{x}_{2,2}\}$. Construir proyectores sobre dichos subespacios con bases duales.

(c) Verificar en Python que $D(D^\top D)^{-1} D^\top$ coincide con la matriz proyector sobre V_2 encontrada en el punto anterior.

(d) Verificar en Python que $\alpha = (D^\top D)^{-1} D^\top \mathbf{y}_{\text{emp}}$ es un vector cuyas componentes coinciden con los coeficientes de la solución de la regresion dada en la figura 5.

(e) Verificar que la suma de los proyectores sobre cada uno de los elementos de la base $\{\mathbf{x}_{1,0}, \mathbf{x}_{1,1}, \mathbf{x}_{2,2}\}$ es la identidad.

Resolverlo antes de ver la solución.

Solución de parte del punto 2: ver figura 6.

```python
# Proyectores con bases duales
Proj_x10 = np.dot(x10, np.transpose(x10d))
Proj_x11 = np.dot(x11, np.transpose(x11d))
Proj_x22 = np.dot(x22, np.transpose(x22d))
```

```python
# Proyector sobre V1 con formula
DTD = np.dot(np.transpose(D),D)
DTD_inv = np.linalg.inv(DTD)
Proj_V1 = np.dot(D,np.dot(DTD_inv,np.transpose(D)))
print(Proj_V1)
# Confirmacion de equivalencia
print(Proj_x10 + Proj_x11)
```

```
[[ 0.83333333  0.33333333 -0.16666667]
 [ 0.33333333  0.33333333  0.33333333]
 [-0.16666667  0.33333333  0.83333333]]
[[ 0.83333333  0.33333333 -0.16666667]
 [ 0.33333333  0.33333333  0.33333333]
 [-0.16666667  0.33333333  0.83333333]]
```

```python
# La identidad como suma de proyectores
print(Proj_x10 + Proj_x11 + Proj_x22)
```

```
[[ 1.00000000e+00  0.00000000e+00 -5.55111512e-17]
 [ 0.00000000e+00  1.00000000e+00 -5.55111512e-17]
 [ 5.55111512e-17  1.11022302e-16  1.00000000e+00]]
```

Figure 6: Solución del ejercicio 4.1.1.

4.2 Condiciones para que el problema de regresión tenga solución

La solución (3.48) al problema de regresión involucra la inversa de la matriz $D^\top D$, donde la matriz D viene dada en (4.2). Es razonable preguntarse sobre las condiciones en las que dicha inversa existe.

Recordemos que la matriz $D \in \mathbb{R}^{n\times j}$, por lo que la matriz $G = D^\top D \in \mathbb{R}^{d\times d}$ (a veces llamada **matriz de Gram**) es simétrica $G^\top = G$ y "semidefinida positiva".

Que es simétrica se ve así: $G^\top = (D^\top D)^\top = D^\top (D^\top)^\top = D^\top D = G$. Donde usamos la propiedad vista en la sección 1.4 que la transpuesta de un producto es el producto de las transpuestas multiplicadas en el orden inverso.

Una matriz cuadrada $G \in \mathbb{R}^{d\times d}$ es semidefinida positiva si para todo vector $\boldsymbol{\alpha} \in \mathbb{R}^{d\times 1}$, $\boldsymbol{\alpha}^\top G \boldsymbol{\alpha} \geq 0$. Que $G = D^\top D$ cumple con esta condición es muy fácil de ver, ya que para todo vector $\boldsymbol{\alpha}$, el cuadrado de la norma ℓ_2 del vector $\mathbf{y} = D\boldsymbol{\alpha} \in \mathbb{R}^{n\times 1}$ es $||\mathbf{y}||_2^2 = \mathbf{y}^\top \mathbf{y} = \boldsymbol{\alpha}^\top D^\top D \boldsymbol{\alpha} = \boldsymbol{\alpha}^\top G \boldsymbol{\alpha}$. Y como vimos en la sección 1.6, $||\mathbf{y}|| \geq 0$, e $||\mathbf{y}|| = 0 \Leftrightarrow \mathbf{y} = \mathbf{0}$.

Recordemos cómo esta construida la matriz D en (4.2). Tenemos una colección de n muestras empíricas, donde cada muestra contiene los $j-1$ valores de las variables $x_1, \ldots, x_{j-1}$ y un valor y_{emp}. En un situación realista de implementación de un modelo de regresión, es de esperar que el número de muestras sea muy superior al número de parámetros, $n \gg j$, y que los datos son recogidos de manera estadísticamente *independiente*.

Estas condiciones implican que con una altísima probabilidad, los j vectores $\mathbf{x}_{1,i} \in R^n$, $i = 0, 1, \ldots, j-1$ definidos en (4.7) (recordemos que el vector $\mathbf{x}_{1,0} = (1, \ldots, 1)^\top$) van a ser linealmente independientes.

Entonces, para parámetros α_i, $i = 0, \ldots, j-1$, no todos cero, cualquier combinación lineal de ellos $\mathbf{y}_{\text{pred}} = D\boldsymbol{\alpha}$ en (4.6) va a ser $\neq \mathbf{0}$. Por lo tanto el cuadrado de su norma $||\mathbf{y}_{\text{pred}}||_2^2 = \mathbf{y}_{\text{pred}}^\top \mathbf{y}_{\text{pred}} = \boldsymbol{\alpha}^\top D^\top D \boldsymbol{\alpha} = \boldsymbol{\alpha}^\top G \boldsymbol{\alpha} > 0$. Es decir, G es *definida* positiva. Esto significa que su espacio nulo $\mathcal{N}(G) = \{\mathbf{0}\}$, por lo que en las condiciones en que normalmente usamos modelos de regresión, G tiene inversa y la solución del problema de regresión existe y es único.✓

Este análisis nos sirve para entender cuándo y por qué el modelo de regresión pudiera estar indeterminado. Como vimos, si las columnas de D son linealmente independientes, cualquier combinación lineal de ellas con parámetros no todos nulos va a ser $\neq \mathbf{0}$. Lo cual implica que G tiene inversa. Sólo si dichas columnas fueran linealmente dependientes entre sí pudiera una combinación lineal de ellas con parámetros no todos nulos ser $= \mathbf{0}$. Veamos algunas situaciones en las que esto pudiera ocurrir:

1. El modelo tiene más o igual parámetros α_i que datos ($j \geq n$).

2. Hay al menos dos variables explicativas x_i con una alta correlación.

3. Los datos no son obtenidos de manera estadísticamente independiente.

1. Si el modelo tiene más o igual parámetros que datos ($j \geq n$), las j columnas van a spanear todo el espacio $\mathbb{R}^n$, es decir, $V_1 = V$. Entonces, en (4.8), $\mathbf{err} = \mathbf{y}_{\text{emp}} - \mathbf{y}_{\text{pred}} = \mathbf{0}$. Si $j = n$, la solución $\{\alpha_i\}$ va a ser única. Si $j > n$, van a haber infinitos conjuntos de parámetros $\{\alpha_i\}$ correspondientes al óptimo, ya que los vectores columna de D van a ser linealmente dependientes. Como sabemos, cualquier vector de V, en particular, $\mathbf{y}_{\text{emp}}$, se puede escribir de infinitas maneras como combinación lineal de bases linealmente dependientes. Como vamos a ver, esto conduce a "overfitting" y el modelo no va a tener poder predictivo.

2. Supongamos que $j \ll n$ pero hay dos variables explicativas, x_r y x_s, con correlación 1 o -1 entre ellas. En ese caso, en la matriz D en (4.2), $x_{i;1,r} = c x_{i;1,s}$ para todo $i = 0, \ldots, n-1$, donde $c \in \mathbb{R}$. Si $c > 0$ la correlación es 1, si $c < 0$ la correlación es -1. En ambos casos, las columnas r y s serán linealmente dependientes. Este problema es en general menos grave que el anterior, porque simplemente requiere que reduzcamos el número de variables explicativas, eliminando las redundantes.

Aún si la correlación no es exactamente ± 1, pero lo suficientemente cercana a 1 en valor absoluto, teniendo en cuenta que los datos empíricos son medidos siempre con cierto error, las variables van a ser redundantes.

Por supuesto todo esto se extiende a situaciones en las que una variable es linealmente dependiente de varias otras.

3. Para entender qué pasa si los datos no son obtenidos de manera estadísticamente independiente, imaginemos una situación en la que de los n samples empíricos, aunque $n \gg j$, muchos son repetidos, por lo que sólo hay $r < j$ datos diferentes. De las n filas de la matriz D, sólo r de ellas van a ser linealmente independientes. Y como el rango fila de una matriz es igual al rango columna, sólo r columnas van a ser linealmente independientes. Por lo que efectivamente estamos en la situación descripta en **1**.

5 Conclusiones

En este trabajo presentamos una primera revision de álgebra linea de una forma especialmente adaptada para sus eventuales aplicaciones en aprendizaje automático (machine learning).